On the Dual Approach to Recursive Optimization

Matthias Messner[*] Nicola Pavoni[†] Christopher Sleet[‡]

October 26, 2011

Abstract

We bring together the theories of duality and dynamic programming. We show that the dual of an additively separable dynamic optimization problem can be recursively decomposed using summaries of past Lagrange multipliers as state variables. Analogous to the Bellman decomposition of the primal problem, we prove equality of values and solution sets for recursive and sequential dual problems. In non-additively separable settings, the equivalence of the recursive and sequential dual is not guaranteed.

We relate recursive dual and recursive primal problems. If the Lagrangian associated with a constrained optimization problem admits a saddle then, even in non-additively separable settings, the values of the recursive dual and recursive primal problems are equal. Additionally, the recursive dual method delivers necessary conditions for a primal optimum. If the problem is strictly concave, the recursive dual method delivers necessary and sufficient conditions for a primal optimum. When a saddle exists, states on the optimal dual path are subdifferentials of the primal value function evaluated at states on the optimal primal path and vice versa.

1 Introduction

Many dynamic economic optimization problems have a recursive structure and may be solved via the application of dynamic programming techniques. The use of these techniques has, consequently, become very widespread in economics. Dynamic incentive

[*]Bocconi University and IGIER, Via Roentgen 1, 20136 Milano, Italy; matthias.messner@unibocconi.it.

[†]Department of Economics and IGIER, Bocconi University, I-20136, Milan, Italy; Department of Economics, EUI, I-50133 Firenze, Italy; CEPR and IFS, London; pavoni.nicola@gmail.com.

[‡]Tepper School of Business, Carnegie Mellon University, Pittsburgh PA 15217; csleet@andrew.cmu.edu.

problems, an important class that includes optimal risk sharing, Ramsey taxation and dynamic principal-agent problems, present special challenges for the dynamic programming approach. First, although these problems often admit a recursive formulation it is typically implicit. The elements of such a formulation, in particular, the state space and constraint correspondence, must be recovered from the primitives of the problem. Second, some candidate states may not be feasible. The "effective" state space of feasible states is endogenous and must be derived as part of the solution to the problem. Third, constraints often run across current shock realizations or involve expectations of future variables. This complicates the recursive decomposition by raising the dimensionality of state and/or choice variables.

The difficulties described above have prompted economists to adopt recursive methods that replace or supplement standard "primal" state variables with "dual" ones. Examples include, inter alia, Kydland and Prescott (1980), Kehoe and Perri (2002), Marimon and Quadrini (2006), Acemoğlu, Golosov, and Tsyvinski (2010), Chien, Cole, and Lustig (2011) and Aiyagari, Marcet, Sargent, and Seppälä (2002). Despite their widespread use, thorough analysis of these methods is limited and their application has often been ad hoc. In this paper, we systematically develop a new recursive dual approach to dynamic optimization. This approach blends elements of the theories of duality and dynamic programming. Recall that a dynamic or sequential (primal) optimization may be formulated in terms of a Lagrangian as $\sup_a \inf_\lambda \mathscr{L}(a;\lambda)$, where a stands for a sequence of control variables and λ for a sequence of Lagrange multipliers attached to constraints. The order of the supremum and infimum operations is important. By interchanging them the dual problem $\inf_\lambda \sup_a \mathscr{L}(a;\lambda)$ is obtained. Our *dual recursive method* (DRM) pairs a primal problem with its dual and then seeks a recursive formulation of the latter. Our approach is in certain respects simpler than the well known *recursive saddle point method* of Marcet and Marimon (1999) (revised: Marcet and Marimon (2011)) and improves upon it by giving optimal primal values and necessary conditions for optimal primal solutions in a larger class of problems.

Three main sets of results underpin our approach. The first relates recursive dual to sequential dual problems; the second relates sequential dual to sequential primal problems. In combination these results tie recursive dual to sequential primal problems. Our method uses the former to solve the latter. The third set of results elaborates the connections between recursive primal and dual problems and gives perspective on the value of the recursive dual approach.

1. *Recursive decomposition of the sequential dual.* Recall that appropriate separability

assumptions on the objective and constraint functions of a sequential (primal) optimization allow it to be recursively decomposed and, hence, solved via the application of dynamic programming techniques. Under these assumptions, the optimal values from the sequential and recursive problems coincide and the solutions generated by the recursive problem's policy functions solve the sequential problem. The most common form of separability found in economic applications is (intertemporal and interstate) additive separability. However, the approach can be extended to accommodate other, weaker forms of separability.

We show that sequential dual problems also admit recursive decompositions and can be solved with dynamic programming whenever the underlying objective and constraint functions are additively separable. Thus, recursive and sequential dual problems with the same optimal values and solution sets are available in many standard economic settings.[1] However, outside of the additive case, equivalence between sequential and recursive dual problems breaks down. The reason is subtle. In moving from the sequential to the recursive dual, a conditional infimum and a supremum operation must be interchanged. The former involves a future multiplier, the latter a current primal variable. If the objective, in this case a "conditional" Lagrangian, allows the infimum and supremum operations to be decoupled, then optimal values and solutions are preserved by the interchange. Additive separability is sufficient for this decoupling, weaker separability assumptions are not. The inf-sup interchange preserves optimal values in non-additively separable problems *if* the conditional Lagrangian objective admits a saddle point.[2] In this case, the recursive and sequential dual problems give the same optimal value and the latter gives necessary conditions for the solutions of the former.

2. *Relationship between the sequential problems*. The relationship between the sequential dual and primal also involves an interchange of infimum and supremum operations. In this case, all inf-sup operations associated with the sequential problem's Lagrangian are involved in the interchange (and not just a component of them). If the sequential Lagrangian admits a saddle point, then the optimal values from the sequential dual and primal coincide and the sequential dual gives necessary, but not sufficient, conditions for an optimal primal solution. Combining results, the recursive dual also gives the optimal primal value and necessary conditions for an optimal primal solution. We emphasize that this relation between the recursive

[1] For example, many settings considered in Stokey, Lucas, and Prescott (1989).

[2] If the sequential dual problem admits a solution, then additive separability ensures the existence of such conditional saddle points.

dual and the primal problem requires the existence of a saddle *only* in the sequential problem and *not* in all sub-problems and after all histories. The latter condition is used in other analyses and is much stronger. Existence of a saddle point does not guarantee that the dual problem delivers sufficient conditions for optimality.[3] Sufficiency can be obtained by imposing stronger concavity assumptions on the optimization problem.

3. *Duality between recursive problems.* Our third set of results identifies additional relations between primal and dual value functions and state variables. We show that if the sequential Lagrangian admits a saddle, then along the optimal path, dual (resp. primal) state variables are sub-differentials of the primal (resp. dual) value function. In addition, under the same assumption, modulo sign changes and again on the optimal path, the dual value function equals the conjugate of the primal value function.[4] To the best of our knowledge no comparable results have been shown in the existing literature. The duality relations that we prove, are useful for understanding the practical merits of the recursive dual and primal approaches. The two approaches place the problem on different state spaces. In many dynamic contracting settings, the natural primal state space is an awkward endogenous set, but the dual state space (the set of primal sub-differentials) is a simple set that is easily predetermined.

   In addition, we note a further duality between primal and dual state variables. It is often useful to classify state variables as "backward-looking" or "forward-looking", where the former are given as functions of an initial state and past controls and the latter are given as functions of a terminal state and future controls. We show that the dual counterparts of backward-looking primal state variables are forward-looking and vice versa.

We pursue our main ideas in an abstract, but simple two period setting that encompasses many concrete economic applications. The restriction to two periods allows us to express our key duality results in a transparent way, while avoiding technicalities and the more complicated notation that arises in infinite horizon settings.[5]

[3] Messner and Pavoni (2004) show that also the recursive saddle point method proposed by Marcet and Marimon (1999) suffers from the same lack of sufficiency even when the sequential Lagrangian admits a saddle. For a more detailed discussion of the relationship between our method and the recursive saddle point method please see Section 7

[4] The conjugate of a function $f : \mathbb{R}^n \to \overline{R}$ is given by $f^\star : \mathbb{R}^n \to \overline{\mathbb{R}}$, where $f^\star(x^\star) = \sup_{\mathbb{R}^n} \langle x, x^* \rangle - f(x)$.

[5] The technicalities relate to the choice of dual and candidate value function spaces. See Messner, Pavoni, and Sleet (2011) for further analysis of infinite period problems.

The paper proceeds as follows. After a brief review of the literature, Section 2 summarizes basic notions of duality and dynamic programming. Section 3 is the core of the paper. In Subsection 3.1, a two period additive framework is laid out that accommodates many economic problems. In this context, a primal sequential optimization is defined. Subsection 3.2 shows how primal states are determined, provides a recursive primal formulation and establishes a primal Bellman-type principle of optimality. Sequential and recursive dual problems are introduced in Subsection 3.3 and a dual version of Bellman's optimality principle is given that relates them. The relation between the sequential primal and sequential dual problems and, hence, between the sequential primal and recursive dual problems is developed in Subsections 3.3 and 3.4. Section 4 extends the analysis to cover families of problems parameterized by initial and terminal conditions for states, while Section 5 gives applications to economic growth and contracting with limited commitment. Section 6 turns to non-additively separable problems and does so in the context of a concrete dynamic moral hazard application. Section 7 gives a detailed comparison of our approach with that of Marcet and Marimon (1999).

Literature In a seminal contribution, Marcet and Marimon (1999) (revised: Marcet and Marimon (2011)) proposed solving dynamic optimizations by recursively decomposing the saddle operation: $\text{saddle}_{(a|\lambda)} \mathscr{L}(a,\lambda) = \mathscr{L}(a^*,\lambda^*)$, with $a^* = \text{argmax}_a \mathscr{L}(a,\lambda^*)$ and $\lambda^* = \text{argmin}_\lambda \mathscr{L}(a^*,\lambda)$. Under this operation, the minimization and maximization steps are done in parallel rather than in sequence making the decomposition of the saddle less direct and more delicate than the decomposition of the dual. As a consequence, Marcet and Marimon (1999) need to impose stronger assumptions in order to derive their central results, than we have to under our approach. More specifically, the main differences between our paper and theirs can be summarized as follows:

i) In order for the recursive saddle point method to be well posed, the optimization problem has to admit a saddle point after every history. The recursive dual problem is well defined independently of any saddle point considerations. Moreover, none of our results require more than the existence of a saddle at the *initial* history.

ii) We obtain necessity (every solution of the underlying optimization problem can be generated by the policy of the recursive dual problem) without imposing any assumptions beyond existence of an initial saddle. On top of the stronger saddle requirements, Marcet and Marimon (2011) also assume that the underlying optimization problem has a unique solution in order to prove necessity.[6]

[6]See Corollary 3.1 of Marcet and Marimon (2011).

iii) They use a mixture of primal and dual state variables, whereas we use one or the other. The use of mixed state variables allows them to handle some non-additive separabilities. We analyze dynamic moral hazard problems, an important class of non-additively separable contracting problem not considered by Marcet and Marimon (1999), and show how they can be handled via the recursive dual approach.[7]

Section 7 further elaborates the distinctions between the recursive saddle and dual approaches.

Messner, Pavoni, and Sleet (2011) develop a recursive approach related to that considered here. In contrast to this paper, in which the primal problem is first "dualized" and then "recursivized", in Messner, Pavoni, and Sleet (2011) these roles are reversed. A recursive primal problem is paired with a recursive dual problem. Messner, Pavoni, and Sleet (2011) focus on infinite horizon problems and establish a duality between primal and dual value iteration. Cole and Kubler (2010) show how dual variables may be used to summarize histories in a recursive primal problem (in which "sup-inf" operations are performed at each step). They show how such a method may be extended to give sufficient conditions for an optimal primal solution without making strict concavity assumptions as are needed here and in Marcet and Marimon (1999).

2 Background Theory

This section collects basic results from the theories of duality and dynamic programming. These results underpin the subsequent analysis. Our account draws mainly on Rockafellar (1970) and Borwein and Lewis (2006).

2.1 Duality

The Lagrangian and the primal problem. Consider the optimization:

$$\sup f(a) \tag{1a}$$

$$\text{s.t. } a \in A \text{ and } g(a) \geq 0, \tag{1b}$$

where $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}^m$. Let V^* denote the optimal value from problem (1) and A^* its solution set. We associate the Lagrangian $\mathscr{L} : A \times \Lambda \to \mathbb{R}$, with

[7]However, see Mele (2010) for an heuristic attempt to extend the recursive saddle point method to dynamic moral hazard settings.

the problem, where $\Lambda := \mathbb{R}^m_+$,

$$\mathscr{L}(a,\lambda) := f(a) + \lambda \cdot g(a), \tag{2}$$

and $\lambda \cdot g(a)$ is the dot product of λ and $g(a)$. Since

$$\inf_{\Lambda} \mathscr{L}(a,\lambda) = \begin{cases} f(a) & \text{if } g(a) \geq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

it follows that (1) is equivalent to the following 'infinite penalization' problem:

$$\sup_{A} \inf_{\Lambda} \mathscr{L}(a,\lambda). \tag{3}$$

Specifically, $A^* = A_{SI} := \operatorname{argmax}_A \inf_\Lambda \mathscr{L}(a,\lambda)$ (SI='sup inf') and $V^* = \sup_A \inf_\Lambda \mathscr{L}(a,\lambda)$.

The Lagrangian and the dual problem. The 'dual' problem to (3) is defined by interchanging the supremum and infimum operations:

$$\inf_{\Lambda} \sup_{A} \mathscr{L}(a,\lambda). \tag{4}$$

Denote the optimal value of the dual by W^* and its solution set by $\Lambda_{IS} := \operatorname{argmin}_\Lambda \sup_A \mathscr{L}(a,\lambda)$ (IS='inf sup'). In addition, for each $\lambda \in \Lambda$, let $A_{IS}(\lambda) := \operatorname{argmax}_A \mathscr{L}(a,\lambda)$ be the set of conditional maximizers and let $A_{IS} := \cup_{\lambda \in \Lambda_{IS}} A_{IS}(\lambda)$.

The ordering of the sup and inf operations matters. By classical weak duality, the dual value W^* weakly exceeds the primal value V^*.[8] However, it may strictly exceed it in which case there is said to be a *duality gap*. In such situations, the dual problem is of limited use in characterizing the primal; the dual solution set A_{IS} may even be disjoint from $A^*(= A_{SI})$. Existence of a *saddle point* for the Lagrangian provides a sufficient condition for the absence of a duality gap.

Definition 1 (Saddle Point)**.** A point $(\hat{a},\hat{\lambda}) \in A \times \Lambda$ is a saddle point of the Lagrangian $\mathscr{L} : A \times \Lambda \to \mathbb{R}$ if:

$$\mathscr{L}(a,\hat{\lambda}) \leq \mathscr{L}(\hat{a},\hat{\lambda}) \leq \mathscr{L}(\hat{a},\lambda) \qquad \text{for all } a \in A,\ \lambda \in \Lambda. \tag{5}$$

Proposition 1. *A pair $(\hat{a},\hat{\lambda})$ is a saddle point of $\mathscr{L}$ if and only if $\hat{a} \in A_{SI}$, $\hat{\lambda} \in \Lambda_{IS}$, and*

[8] See, for example, Rockafellar (1970).

$V^* = W^* = \mathscr{L}(\hat{a}, \hat{\lambda})$. *Also, if* $\mathscr{L}$ *admits a saddle point, then* $A^* = A_{SI} \subset A_{IS}(\hat{\lambda})$ *for each* $\hat{\lambda} \in \Lambda_{IS}$.

Proof. See, for example, Rockafellar (1970). □

Proposition 1 implies that if a saddle exists, then the dual and primal problems have the same optimal value and the dual problem gives necessary conditions for the primal solution set A^*. Specifically, in this case, for any multiplier $\hat{\lambda}$ that attains the minimum in the dual, all primal solutions $\hat{a} \in A^*$ belong to $A_{IS}(\hat{\lambda})$ and solve the maximization $\max_A \mathscr{L}(a, \hat{\lambda})$. However, absent furher assumptions we cannot rule out that the inclusion $A^* \subset A_{IS}(\hat{\lambda})$ is strict and that $A_{IS}(\hat{\lambda})$ contains elements that do not solve (and may not even be feasible for) the primal problem. Thus, even if a saddle exists, the dual problem (4) may not give sufficient conditions for a primal solution.

Optimization and Conjugate Duality. Relations between primal and dual optimizations can be expressed in terms of conjugate functions. The (Fenchel) *conjugate* of $f : \mathbb{R}^n \to [-\infty, \infty]$ is given by $f^\star : \mathbb{R}^n \to [-\infty, \infty]$, where:

$$f^\star(\lambda) := \sup_{\mathbb{R}^n} \{\lambda \cdot x - f(x)\}. \tag{6}$$

The function $f^{\star\star} = (f^\star)^\star$ is called the *biconjugate* of f. The sub-differential of a function $f : \mathbb{R}^n \to \mathbb{R}$ at the point $\bar{x}$ is given by:

$$\partial f(\bar{x}) = \{\lambda \in \mathbb{R}^n | \forall x \in \mathbb{R}^n, \quad f(x) \geq f(\bar{x}) + \lambda \cdot (x - \bar{x})\}.$$

In the paper, we make use of the following duality between the sub-differentials of functions and their conjugates.

Proposition 2. *Let* $f : \mathbb{R}^n \to [-\infty, \infty]$. *If* $\bar{\lambda} \in \partial f(\bar{x})$ *then* $\bar{x} \in \partial f^\star(\bar{\lambda})$.

Proof. See Rockafellar (1970). □

Consider the parametrized problem,

$$\sup_A f(a) \qquad \text{s.t. } g(a) \geq b,$$

where $b \in \mathbb{R}^m$. Let $-P(b)$ denote the optimal value for this problem so that $V^* = -P(0)$. In addition, let $D(\lambda) = \sup_A \mathscr{L}(a, \lambda)$.

Proposition 3. *The conjugate of the function P satisfies:*

$$P^{\star}(\lambda) = \begin{cases} D(\lambda) & \text{if } \lambda \in \Lambda; \\ +\infty & \text{otherwise.} \end{cases}$$

Also, $-W^* = P^{\star\star}(0)$. Hence, $W^* = V^*$ and there is no duality gap if and only if $P(0) = P^{\star\star}(0)$.

Proof. See Rockafellar (1970). □

Notice that Proposition 3 does not use convexity. Since the biconjugate of a function is its convex, lower semicontinuous regularization, conditions that ensure the convexity and lower semicontinuity of P globally are sufficient, but not necessary for primal and dual values to coincide. Our final result links the existence of saddle points to the subdifferentiability of the value function P.

Proposition 4. *Suppose the optimization problem (1) has a solution and, hence, a finite value V^*. Then the Lagrangian $\mathscr{L}$ admits a saddle point if and only if $\partial P(0)$ is non-empty. In particular, for any given $\hat{a} \in A^*$, $\hat{\lambda} \in \Lambda$ solves $\inf_{\lambda} \mathscr{L}(\hat{a}, \lambda)$ if and only if $\hat{\lambda} \in \partial P(0)$.*

Proof. See Rockafellar (1970). □

Again Proposition 4 does not require explicit convexity assumptions.

2.2 Pre-Bellman Principle of Optimality

Duality theory handles mixed combinations of supremum and infimum operations. It implies that only under specific conditions are optimal values and solutions preserved by the interchange of these operations. On the other hand, as the following result shows, optimal values and solutions are unaffected by the interchange and recombination of the same type of operation.[9]

Proposition 5 (Pre-Bellman Optimality Principle)**.** *Let $f : B \to \mathbb{R}$, where $B \subset A_1 \times A_2$, then:*

$$\sup_{(a_1,a_2)\in B} f(a_1,a_2) = \sup_{a_1\in B_1} \left\{ \sup_{a_2\in B_2(a_1)} f(a_1,a_2) \right\} = \sup_{a_2\in B_2} \left\{ \sup_{a_1\in B_1(a_2)} f(a_1,a_2) \right\},$$

[9] We give the supremum version, evidently an infimum version is available.

where $B_1 = \{a_1 \in A_1 | \exists a_2$ *with* $(a_1, a_2) \in B\}$ *and* $B_1(a_2) = \{a_1 \in A_1 | (a_1, a_2) \in B\}$ *and similarly for* B_2 *and* $B_2(a_1)$*. If* (a_1^*, a^*) *attains the supremum (or suprema) in any one of the above problems, then it does in the other two.*

Given the elementary nature of the result, we omit the proof. Note that Proposition 5 does not assume and makes no use of any separability in the objective and constraints. More refined versions of the result that do are given later in the paper.

3 The Recursive Dual Approach in Additive Settings

This section introduces the recursive dual approach to optimization. It does so in an additively separable setting that accommodates many economic applications. Extensions of our method to environments featuring weaker forms of separability are considered later in Section 6. The main conceptual ideas underpinning the recursive dual method are fully captured in a two period framework. To make the analysis and exposition as transparent as possible we adopt such a framework.[10]

3.1 The Additive Model

We impose an additive structure on the problem:

$$\sup\{f(a) | a \in A \text{ and } g(a) \geq 0\}. \tag{7}$$

First, assume that A is a non-empty product set with $A = A_1 \times A_2$ and $A_t \subset \mathbb{R}^{n_t}$, $t = 1, 2$. We interpret t as a time index and problem (7) as dynamic. Second, let the objective function be given by $f : A \to \mathbb{R}$, where $f(a_1, a_2) = f_1(a_1) + f_2(a_2)$ and $f_t : A_t \to \mathbb{R}$. Let the constraint function $g : A \to \mathbb{R}^{k_1+k_2}$ satisfy:

$$g(a) = \begin{bmatrix} g_1(a_1) & + & c_1 \cdot m(a_2) \\ c_2 \cdot l(a_1) & + & g_2(a_2) \end{bmatrix}, \tag{8}$$

[10] We also do not make stochastic elements of the model explicit. Insofar as there is separability of functions defining the model across shock realizations, additional opportunities for decomposition and the recursive dual method are created.

with $g_t : A_t \to \mathbb{R}^{k_t}$, $m : A_2 \to \mathbb{R}^{r_1}$ and $l : A_1 \to \mathbb{R}^{r_2}$. The range indices are assumed to satisfy $r_t < \min(k_t, n_t)$ with c_t a matrix of dimension $k_t \times r_t$, $t = 1, 2$. Problem (7) is then:

$$\begin{aligned} V^* = \sup_{A_1 \times A_2} \quad & f_1(a_1) + f_2(a_2) \\ \text{s.t.} \quad & \begin{array}{ll} g_1(a_1) & + c_1 \cdot m(a_2) \geq 0 \\ c_2 \cdot l(a_1) + g_2(a_2) & \geq 0. \end{array} \end{aligned} \tag{9}$$

Henceforth, we refer to (9) as the *sequential primal problem*. We assume throughout that (9) has a non-empty constraint set and associate with it the Lagrangian $\mathscr{L} : A \times \Lambda \to \mathbb{R}$, $\Lambda = \mathbb{R}_+^{k_1} \times \mathbb{R}_+^{k_2}$, where:

$$\mathscr{L}(a, \lambda) := f_1(a_1) + f_2(a_2) + \lambda_1 \cdot [g_1(a_1) + c_1 \cdot m(a_2)] + \lambda_2 \cdot [c_2 \cdot l(a_1) + g_2(a_2)]. \tag{10}$$

From the discussion in Subsection 2.1, the sequential primal problem (9) is equivalent to:

$$V^* = \sup_A \inf_\Lambda \ \mathscr{L}(a, \lambda). \tag{11}$$

3.2 The recursive primal problem.

We present the recursive primal approach in a form that brings out its symmetry with the recursive dual approach. The former approach decomposes the sequential problem (9) into first and second period sub-problems. The decomposition involves breaking the constraints in (8) into components that are expressed in terms of and linked by state variables. Component constraints are allocated to periods and used to define the sub-problems. For practical computational reasons, it is useful to choose a decomposition that minimizes the dimension of the state variables. Since elements of $Y := m(A_2)$ have lower dimension than elements of $g_1(A_1)$ or A_2 and elements of $X := l(A_1)$ have lower dimension than elements of $g_2(A_2)$ or A_1, X and Y are natural candidate state spaces.

Using the state spaces X and Y, the sequential problem (9) can be decomposed as:

$$\begin{aligned} V_1 \quad := \sup_{A_1 \times X \times Y} \quad & f_1(a_1) + V_2(x, y) \\ \text{s.t.} \quad & g_1(a_1) + c_1 \cdot y \geq 0; \ \ x = l(a_1), \end{aligned} \tag{12}$$

where

$$\begin{aligned} V_2(x,y) &:= \sup_{A_2} f_2(a_2) \qquad (13) \\ &\text{s.t.} \quad g_2(a_2) + c_2 \cdot x \geq 0; \;\; y = m(a_2), \end{aligned}$$

and the convention $\sup \emptyset = -\infty$ is used.

Remark 1. The terms "backward" and "forward-looking" state variable are often used informally in economics. We define a state variable as backward-looking if it is given as a function of past choices and forward-looking if it is given as a function of future ones. Thus, x is backward-looking and y forward-looking. In economic contexts, if capital is given as a function of past investments we would label it backward-looking, if a utility promise from a principal to an agent is a function of future rewards to the agent, we would label it forward-looking.[11] $\square$

For the purposes of comparison with the recursive dual method, it is useful to give a restatement of the recursive primal problem in terms of sup-inf operations:

$$\begin{aligned} V_1 &= \sup_{A_1 \times X \times Y} \inf_{\Lambda_1} f_1(a_1) + \lambda_1 \cdot (g_1(a_1) + c_1 \cdot y) + V_2(x,y) \qquad (14) \\ &\text{s.t. } x = l(a_1) \end{aligned}$$

and:

$$\begin{aligned} V_2(x,y) = &\sup_{A_2} \inf_{\Lambda_2} f_2(a_2) + \lambda_2 \cdot (c_2 \cdot x + g_2(a_2)) \\ &\text{s.t. } y = m(a_2), \end{aligned}$$

where $\Lambda_t := \mathbb{R}_+^{k_t}$, $t = 1, 2$. Let $\Gamma_1 \subset A_1 \times X \times Y \times \Lambda_1$, with

$$\Gamma_1 := \left\{ \begin{array}{l} (a_1^*, x^*, y^*, \lambda_1^*) \in A_1 \times X \times Y \times \Lambda_1 : \; V_1 = f_1(a_1^*) + V_2(x^*, y^*), \; x^* = l(a_1^*), \\ \text{and } \lambda_1^* \in \operatorname{argmin}_{\Lambda_1} \; f_1(a_1^*) + \lambda_1 \cdot [g_1(a_1^*) + c_1 \cdot y^*] + V_2(x^*, y^*) \end{array} \right\},$$

[11] However, the labels should be used with care. In some settings a variable could be classified as both forward and backward; in others, the classification of the same economic concept might depend on the exact specification of the problem. In these situations the nomenclature of backwards and forwards may not be very helpful. However, when it is we use it.

and $\Gamma_2 : X \times Y \to 2^{A_2 \times \Lambda_2}$, with

$$\Gamma_2(x,y) := \left\{ \begin{array}{l} (a_2^*, \lambda_2^*) \in A_2 \times \Lambda_2 : \; V_2(x,y) = f_2(a_2^*),\; y = m(a_2^*), \\ \text{and } \lambda_2^* \in \operatorname{argmin}_{\Lambda_2} \; f_2(a_2^*) + \lambda_2 \cdot [c_2 \cdot x + g_2(a_2^*)] \end{array} \right\},$$

denote the optimal policy set and correspondence associated with (14). Note that our definition of optimal policies includes the minimizing Lagrange multipliers λ_1^* and λ_2^*. Their inclusion again emphasizes the symmetry between the recursive primal and dual approaches and, as we show below, is essential for the latter. The equivalence of the recursive and sequential primal problems is proved next. Recall that we denote the solution set of the sequential primal problem (11) by A_{SI} and the set of Lagrange multipliers that solve $\min_\Lambda \mathscr{L}(a^*, \lambda)$, $a^* \in A_{SI}$, by $\Lambda_{SI}(a^*)$.

Proposition 6 (Bellman (1957)). *1.* $V_1 = V^*$.

2. *If* $(a_1^*, a_2^*) \in A_{SI}$ *and* $\lambda^* \in \Lambda_{SI}(a^*)$, *then there exist* $x^* \in X$ *and* $y^* \in Y$ *such that* $(a_1^*, x^*, y^*, \lambda_1^*) \in \Gamma_1$ *and* $(a_2^*, \lambda_2^*) \in \Gamma_2(x^*, y^*)$. *Conversely, if* $(a_1^*, x^*, y^*, \lambda_1^*) \in \Gamma_1$ *and* $(a_2^*, \lambda_2^*) \in \Gamma_2(x^*, y^*)$, *then* $(a_1^*, a_2^*) \in A_{SI}$ *and* $(\lambda_1^*, \lambda_2^*) \in \Lambda_{SI}(a_1^*, a_2^*)$.

Proof. See Appendix. □

The proposition asserts that the sequential and recursive primal values coincide, that any selection from the recursive policy correspondence solves the sequential problem and that any solution to the sequential problem corresponds to a selection from the recursive policy correspondence.

3.3 The recursive dual approach

The *sequential dual problem* associated with (9) interchanges the sup and inf operations. It is:

$$\begin{aligned} W^* &:= \inf_{\Lambda} \sup_{A} \mathscr{L}(a, \lambda) \\ &= \inf_{\Lambda} \sup_{A} f_1(a_1) + f_2(a_2) + \lambda_1[g_1(a_1) + c_1 \cdot m(a_2)] + \lambda_2[c_2 \cdot l(a_1) + g_2(a_2)]. \end{aligned} \tag{15}$$

By rearranging the Lagrangian terms involving first period choices a_1 can be separated from those involving future choices a_2:[12]

$$W^* = \inf_{\Lambda} \sup_{A} f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \lambda_2 \cdot c_2 \cdot l(a_1) + [f_2(a_2) + \lambda_1 \cdot c_1 \cdot m(a_2) + \lambda_2 \cdot g_2(a_2)].$$

To proceed with the recursive decomposition of (15), we introduce a pair of *dual state variables*, $\mu \in M := \{\lambda_1 \cdot c_1 | \lambda_1 \in \Lambda_1\} \subset \mathbb{R}^{r_1}$ and $\phi \in F := \{\lambda_2 \cdot c_2 | \lambda_2 \in \Lambda_2\} \subset \mathbb{R}^{r_2}$. μ summarizes aspects of the first period multiplier λ_1 that are relevant for the second period problem (i.e. the weight $\lambda_1 \cdot c_1$ on $m(a_2)$), while ϕ summarizes aspects of the second period multiplier λ_2 that are relevant for the first period problem (i.e. the weight $\lambda_2 \cdot c_2$ on $l(a_1)$). Since μ summarizes the multiplier history and ϕ the future multiplier path, we refer to them as backward and forward-looking respectively.

Using these states the sequential dual problem (15) can be decomposed as follows.

Proposition 7. *Let:*

$$\begin{aligned} W_1 \quad &:= \inf_{\Lambda_1 \times M \times F} \sup_{A_1} f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \phi \cdot l(a_1) + W_2(\mu, \phi) \\ &\quad s.t. \quad \mu = \lambda_1 \cdot c_1. \end{aligned} \tag{16}$$

where $W_2 : M \times F \to \overline{\mathbb{R}}$ *is given by:*

$$\begin{aligned} W_2(\mu, \phi) := &\inf_{\Lambda_2} \sup_{A_2} f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \\ &s.t. \quad \phi = \lambda_2 \cdot c_2. \end{aligned} \tag{17}$$

Then $W^* = W_1$.

Proof. Define $\mathrm{Gr}\, M = \{(\lambda_1, \mu) | \lambda_1 \in \Lambda_1, \mu = \lambda_1 \cdot c_1\}$ and $\Lambda_2(\phi) = \{\lambda_2 \in \Lambda_2 : \phi = \lambda_2 \cdot c_2\}$.

[12] This rearrangement is innocuous in the present finite dimensional setting. In infinite dimensional settings additional 'regularity' conditions are required.

Then:

$$
\begin{aligned}
W^* &= \inf_{\Lambda} \sup_{A} \left\{ f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \lambda_2 \cdot c_2 \cdot l(a_1) + \left[f_2(a_2) + \lambda_1 \cdot c_1 \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right] \right\} \\
&= \inf_{\mathrm{Gr}\, M \times F} \inf_{\Lambda_2(\phi)} \sup_{A} \left\{ f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \phi \cdot l(a_1) + f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right\} \\
&= \inf_{\mathrm{Gr}\, M \times F} \inf_{\Lambda_2(\phi)} \sup_{A_1} \Bigg\{ f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \phi \cdot l(a_1) \\
&\qquad\qquad + \sup_{A_2} \left\{ f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right\} \Bigg\} \\
&= \inf_{\mathrm{Gr}\, M \times F} \sup_{A_1} \Bigg\{ f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \phi \cdot l(a_1) \\
&\qquad\qquad + \inf_{\Lambda_2(\phi)} \sup_{A_2} \left\{ f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right\} \Bigg\} \\
&= \inf_{\mathrm{Gr}\, M \times F} \sup_{A_1} \left\{ f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \lambda_2 \cdot l(a_1) + W_2(\mu, \phi) \right\} = W_1.
\end{aligned}
$$

The first equality follows from the definition of W^*, the second from the definitions of Gr M, μ, ϕ and the decomposition properties of the infimum operation and the third from the decomposition properties of the supremum operation (see Proposition 5). The fourth row interchanges the infimum operation over λ_2 conditional on ϕ with the supremum operation over a_1. The fourth equality stems from the fact that the choice of a_1 is independent of λ_2 given ϕ. The fifth equality follows from the definition of W_2 and the sixth from the definition of W_1. □

Remark 2. Proposition 7 is the dual analogue of the first part of Proposition 6. Notice that in the fourth of the sequence of equalities in the proof, a conditional infimum operation over λ_2 is interchanged with a supremum operation over a_1. Defining for all $(a_1, \lambda_2) \in A_1 \times \Lambda_2$, $H(a_1) := f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \phi \cdot l(a_1)$, $Q(\lambda_2) := \sup_{A_2}\{f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2)\}$ and $\mathscr{M}(a_1, \lambda_2) = H(a_1) + Q(\lambda_2)$ and inspecting the proof, we see that this interchange replaces: $\inf_{\Lambda_2(\phi)} \sup_{A_1} \mathscr{M}(a_1, \lambda_2)$ with $\sup_{A_1} \inf_{\Lambda_2(\phi)} \mathscr{M}(a_1, \lambda_2)$. This is permissible because the conditional objective $\mathscr{M}$ is additively separable in a_1 and λ_2 enabling the infimum and supremum operations to be fully decoupled:

$$\inf_{\Lambda_2(\phi)} \sup_{A_1} \mathscr{M}(a_1, \lambda_2) = \sup_{A_1} H(a_1) + \inf_{\Lambda_2(\phi)} Q(\lambda_2) = \sup_{A_1} \inf_{\Lambda_2(\phi)} \mathscr{M}(a_1, \lambda_2).$$

Of course, if $a_1^* \in \operatorname{argmax}_{A_1} H(a_1)$ and $\lambda_2^* \in \operatorname{argmin}_{\Lambda_2(\phi)} Q(\lambda_2)$, then (a_1^*, λ_2^*) is a saddle point of $\mathscr{M}$ on $A_1 \times \Lambda_2(\phi)$. In fact, $a_1^* \in \operatorname{argmax}_{A_1} \mathscr{M}(a_1, \lambda_2)$ for all $\lambda_2 \in \Lambda_2(\phi)$ (and not just for λ_2^*) and $\lambda_2^* \in \operatorname{argmin}_{\Lambda_2(\phi)} \mathscr{M}(a_1, \lambda_2)$ for all $a_1 \in A_1$ (and not just for a_1^*). We do not need an explicit saddle point assumption or convexity conditions on the functions used to construct $\mathscr{M}$ to undertake the inf-sup interchange. □

We define the *dual policy correspondences* as follows:

$$\hat{\Gamma}_1 := \left\{ \begin{array}{l} (\hat{\lambda}_1, \hat{\mu}, \hat{\phi}, \hat{a}_1) \in \Lambda_1 \times M \times F \times A_1 | \\ \qquad W_1 = f_1\left(\hat{a}_1\right) + \hat{\lambda}_1 \cdot g_1\left(\hat{a}_1\right) + \hat{\phi} \cdot l\left(\hat{a}_1\right) + W_2\left(\hat{\mu}, \hat{\phi}\right), \\ \qquad \hat{\mu} = \hat{\lambda}_1 \cdot c_1 \\ \qquad \hat{a}_1 \in \underset{A_1}{\operatorname{argmax}}\ f_1\left(a_1\right) + \hat{\lambda}_1 \cdot g_1\left(a_1\right) + \hat{\phi} \cdot l\left(a_1\right) + W_2\left(\hat{\mu}, \hat{\phi}\right) \end{array} \right\}$$

and

$$\hat{\Gamma}_2\left(\mu, \phi\right) := \left\{ \begin{array}{l} (\hat{\lambda}_2, \hat{a}_2) \in \Lambda_2 \times A_2 | \\ \qquad W_2\left(\mu, \phi\right) = f_2\left(\hat{a}_2\right) + \mu \cdot m\left(\hat{a}_2\right) + \hat{\lambda}_2 \cdot g_2\left(\hat{a}_2\right), \phi = \hat{\lambda}_2 \cdot c_2, \\ \qquad \hat{a}_2 \in \underset{A_2}{\operatorname{argmax}}\ \ f_2\left(a_2\right) + \mu \cdot m\left(a_2\right) + \hat{\lambda}_2 \cdot g_2\left(a_2\right) \end{array} \right\}.$$

Recall that Λ_{IS} is the solution set of the sequential dual problem. For every $\lambda \in \Lambda_{IS}$, let $A_{IS}(\lambda)$ denote the set of controls that solve the problem $\sup_A \mathscr{L}(a, \lambda)$. The next result is the dual analogue of the second part of Proposition 6.

Proposition 8. *If $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2) \in \Lambda_{IS}$ and $(\hat{a}_1, \hat{a}_2) \in A_{IS}(\hat{\lambda})$ then there are values $\hat{\mu}$ and $\hat{\phi}$ such that $(\hat{\lambda}_1, \hat{\mu}, \hat{\phi}, \hat{a}_1) \in \hat{\Gamma}_1$ and $(\hat{\lambda}_2, \hat{a}_2) \in \hat{\Gamma}_2(\hat{\mu}, \hat{\phi})$. Conversely, if $(\hat{\lambda}_1, \hat{\mu}, \hat{\phi}, \hat{a}_1) \in \hat{\Gamma}_1$ and $(\hat{\lambda}_2, \hat{a}_2) \in \hat{\Gamma}_2(\hat{\mu}, \hat{\phi})$, then $\hat{\lambda} \in \Lambda_{IS}$ and $\hat{a} \in A_{IS}(\hat{\lambda})$.*

Proof. Suppose that $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2) \in \Lambda_{IS}$ and $\hat{a} = (\hat{a}_1, \hat{a}_2) \in A_{IS}(\hat{\lambda})$. Let $\hat{\mu} = \hat{\lambda}_1 \cdot c_1$ and $\hat{\phi} = \hat{\lambda}_2 \cdot c_2$. Since $\hat{a} \in A_{IS}$, we have that:

$$\begin{aligned} \hat{a} \in & \underset{A}{\operatorname{argmax}} \left\{ f_1\left(a_1\right) + \hat{\lambda}_1 \cdot g_1\left(a_1\right) + \hat{\lambda}_2 \cdot c_2 \cdot l\left(a_1\right) + f_2\left(a_2\right) + \hat{\lambda}_1 \cdot c_1 \cdot m\left(a_2\right) + \hat{\lambda}_2 \cdot g_2\left(a_2\right) \right\} \\ = & \underset{A}{\operatorname{argmax}}\ \left\{ f_1\left(a_1\right) + \hat{\lambda}_1 \cdot g_1\left(a_1\right) + \hat{\phi} \cdot l\left(a_1\right) + f_2\left(a_2\right) + \hat{\mu} \cdot m\left(a_2\right) + \hat{\lambda}_2 \cdot g_2\left(a_2\right) \right\}. \end{aligned}$$

Since the objective is additively separable in a_1 and a_2, we have, using the definition of W_2,

$$\hat{a}_1 \in \underset{A_1}{\operatorname{argmax}} f_1(a_1) + \hat{\lambda}_1 \cdot g_1(a_1) + \hat{\phi} \cdot l(a_1) + W_2(\hat{\mu}, \hat{\phi})$$

and,

$$\hat{a}_2 \in \underset{A_2}{\operatorname{argmax}} f_2(a_2) + \hat{\mu} \cdot m(a_2) + \hat{\lambda}_2 \cdot g_2(a_2).$$

Also,

$$\begin{aligned}
W_1 &= W^* \\
&= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + f_2(\hat{a}_2) + \hat{\mu} \cdot m(\hat{a}_2) + \hat{\lambda}_2 \cdot g_2(\hat{a}_2) \\
&= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + \sup_{A_2} \left\{ f_2(a_2) + \hat{\mu} \cdot m(a_2) + \hat{\lambda}_2 \cdot g_2(a_2) \right\} \\
&\geq f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + \inf_{\Lambda(\hat{\phi})} \sup_{A_2} \left\{ f_2(a_2) + \hat{\mu} \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right\} \\
&= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + W_2(\hat{\mu}, \hat{\phi}) \\
&= \sup_{A_1} f_1(a_1) + \hat{\lambda}_1 \cdot g_1(a_1) + \hat{\phi} \cdot l(a_1) + W_2(\hat{\mu}, \hat{\phi}) \\
&\geq \inf_{\mathrm{Gr}\, M \times F} \sup_{A_1} f_1(a_1) + \hat{\lambda}_1 \cdot g_1(a_1) + \hat{\phi} \cdot l(a_1) + W_2(\hat{\mu}, \hat{\phi}) = W_1,
\end{aligned}$$

where the first line is from Proposition 7, the second from the definitions of W^*, $\hat{\lambda}$ and $\hat{a}$, the third from the previously proved property of $\hat{a}_2$, the fourth from $\hat{\lambda}_2 \in \Lambda_2(\hat{\phi})$, the fifth from the definition of W_2, the sixth from the previously proved property of $\hat{a}_1$, the seventh from $(\hat{\lambda}_1, \hat{\mu}, \hat{\phi}) \in \mathrm{Gr}M \times F$ and the the definition of W_1. Hence, $W_1 = f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + W_2(\hat{\mu}, \hat{\phi})$ and $(\hat{\lambda}_1, \hat{\phi}, \hat{\mu}, \hat{a}_1) \in \hat{\Gamma}_1$. Also,

$$\begin{aligned}
W^* &= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + f_2(\hat{a}_2) + \hat{\mu} \cdot m(\hat{a}_2) + \hat{\lambda}_2 \cdot g_2(\hat{a}_2) \\
&= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + \inf_{\Lambda(\hat{\phi})} \sup_{A_2} \left\{ f_2(a_2) + \hat{\mu} \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \right\}.
\end{aligned}$$

Subtracting the first from the second line and using the definition of W_2 gives:

$$W_2(\hat{\mu}, \hat{\phi}) = f_2(\hat{a}_2) + \hat{\mu} \cdot m(\hat{a}_2) + \hat{\lambda}_2 \cdot g_2(\hat{a}_2).$$

Hence, $(\hat{\lambda}_2, \hat{a}_2) \in \hat{\Gamma}_2(\hat{\mu}, \hat{\phi})$. It remains to show the converse.

Let $(\hat{a}, \hat{\lambda})$ be a selection from the recursive policy correspondence. Then:

$$
\begin{aligned}
W^* &= W_1 \\
&= f_1(\hat{a}_1) + \hat{\lambda}_1 \cdot g_1(\hat{a}_1) + \hat{\phi} \cdot l(\hat{a}_1) + f_2(\hat{a}_2) + \hat{\mu} \cdot m(\hat{a}_2) + \hat{\lambda}_2 \cdot g_2(\hat{a}_2) \\
&= \sup_{A_1} f_1(a_1) + \hat{\lambda}_1 \cdot g_1(a_1) + \hat{\phi} \cdot l(a_1) + \sup_{A_2} \left\{ f_2(a_2) + \hat{\mu} \cdot m(a_2) + \hat{\lambda}_2 \cdot g_2(a_2) \right\} \\
&= \sup_{A} f_1(a_1) + \hat{\lambda}_1 \cdot g_1(a_1) + \hat{\phi} \cdot l(a_1) + f_2(a_2) + \hat{\mu} \cdot m(a_2) + \hat{\lambda}_2 \cdot g_2(a_2) \\
&\geq \inf_{\Lambda} \sup_{A} f_1(a_1) + \lambda_1 \cdot g_1(a_1) + \lambda_2 \cdot c_1 \cdot l(a_1) + f_2(a_2) + \lambda_1 \cdot c_2 \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \\
&= W^*,
\end{aligned}
$$

where the first equality is from Proposition 7, the second and third from the definitions of $\hat{a}$ and $\hat{\lambda}$ and the policy correspondences, the fourth from Proposition 5, the fifth from the infimum operation and the sixth from the definition of W^*. It follows that $\hat{a} \in A_{IS}(\hat{\lambda})$ and $\hat{\lambda} \in \Lambda_{IS}$. □

Together Propositions 7 and 8 provide a dual version of Bellman's optimality principle for additively separable problems. Combining them with Proposition 1, which relates the sequential dual and primal problems, the following result obtains.

Proposition 9. *(Necessity) If the Lagrangian $\mathscr{L}$ for the sequential (additively separable) problem 9 admits a saddle point, then $W_1 = V^*$, i.e. the recursive dual problem delivers the correct value for the sequential primal problem. In addition, if $\mathscr{L}$ admits a saddle and $\hat{a}$ solves the sequential (primal) problem, then there are values $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu}, \hat{\phi})$ such that $(\hat{\lambda}_1, \hat{\mu}, \hat{\phi}, \hat{a}_1) \in \hat{\Gamma}_1$ and $(\hat{\lambda}_2, \hat{a}_2) \in \hat{\Gamma}_2(\hat{\mu}, \hat{\phi})$.*

Proof. The first part of the proposition is immediate consequence of the two equalities $W_1 = W^*$ and $W^* = V^*$, where the first has been shown in Proposition 7 and the second follows from Proposition 1. The second part of the proposition is an immediate consequence of Propositions 1 and 8. □

Remark 3. The conditions in the previous proposition are weaker than those assumed elsewhere in the literature. In particular, a detailed discussion how our necessity results compared to the one of Marcet and Marimon (1999) is provided in Section 7. □

To illustrate the role of the assumptions and the scope of the results, we give two examples. In the first, there is no saddle and a duality gap emerges: $W^* \neq V^*$. However, since the example is an additively separable one, the recursive and sequential duals remain equivalent.

Example 1. [Lack of Necessity] Consider the following two period problem:

$$V^* = \max_{(a_1,a_2)\in[0,1]^2} \quad -a_1 - a_2 \tag{18}$$

$$\text{s.t.} \quad a_1^2 + a_2^2 \geq b^2 \quad \text{and} \quad a_2^2 \geq c^2, \tag{19}$$

where $c < b < 1$. This problem admits a unique solution, $(a_1^*, a_2^*) = (0, b)$, which generates the value $V^* = -b$.

The dual problem associated with (18) is given by:

$$W^* = \inf_{\lambda_1,\lambda_2\geq 0} \sup_{a_1,a_2\in[0,1]} -a_1 - a_2 + \lambda_1(a_1^2 + a_2^2 - b^2) + \lambda_2(a_2^2 - c^2). \tag{20}$$

It is easy to verify that the unique solution for the multipliers is $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2) = (1, 0)$ and that $A_{IS}(\hat{\lambda}) = \{(0,0),(1,0),(0,1),(1,1)\}$. The optimal dual value is $W^* = -b^2 > V^* = -b$. Hence, there is a duality gap and A_{IS} and A^* are disjoint.

We now verify equivalence of the recursive and sequential duals in this example. Consider the recursive problem:[13]

$$\begin{aligned} W_1 &= \inf_{\lambda_1\in\mathbb{R}_+} \sup_{a_1\in[0,1]} -a_1 + \lambda_1(a_1^2 - b^2) + W_2(\lambda_1), \qquad \text{where} \\ W_2(\lambda_1) &= \inf_{\lambda_2\in\mathbb{R}_+} \sup_{a_2\in[0,1]} -a_2 + \lambda_1 a_2^2 + \lambda_2(a_2^2 - c^2). \end{aligned} \tag{21}$$

It is easily shown that the second period solution is: $\hat{\lambda}_2(\lambda_1) = 1 - \lambda_1$ and $\hat{a}_2 \in \{0, 1\}$. These choices imply:

$$W_2(\lambda_1) = \begin{cases} -(1-\lambda_1)c^2 & \text{if } \lambda_1 < 1 \\ -(1-\lambda_1) & \text{otherwise.} \end{cases}$$

Now consider the first period. The slope of the objective function is negative in λ_1 if $\lambda_1 < 1$ (in which case $a_1 = 0$) and positive for $\lambda_1 > 1$ (in which case $a_1 = 1$). Hence, $\hat{\lambda}_1 = 1$ and $\hat{a}_1 \in \{0, 1\}$. Consequently, the recursive dual approach delivers exactly the same result as the sequential one. $\square$

Proposition 9 shows that if a sequential saddle exists, then the recursive dual method gives necessary conditions for an optimal solution. Our second example shows that even if a saddle exists, the recursive dual method may not give sufficient conditions for

[13] Note in this case $c_1 = 1$, $\lambda_1 = \mu$ and ϕ is superfluous.

a solution.

Example 2. [Lack of Sufficiency] Consider the following two period problem:

$$V^* = \max_{(a_1,a_2)\in[0,1]^2} \quad -a_1 - a_2 \tag{22}$$

$$\text{s.t.} \quad a_1 + a_2 \geq b \quad \text{and} \quad a_2 \geq c, \tag{23}$$

where $c < b < 1$. It is easily seen that the solution set of this problem, A^*, is given by all pairs of non-negative controls (a_1, a_2) which do not violate the second constraint (i.e. $a_2 \geq c$) and satisfy the first constraint with equality (i.e. $a_1 + a_2 = b$). The value of the problem is $V^* = -b$.

Consider the dual problem:

$$W^* = \inf_{\lambda_1,\lambda_2\geq 0} \sup_{a_1,a_2\in[0,1]} -a_1 - a_2 + \lambda_1(a_1 + a_2 - b) + \lambda_2(a_2 - c). \tag{24}$$

The unique solution of this problem is $(\hat{\lambda}_1, \hat{\lambda}_2) = (1, 0)$. For every multiplier pair such that $\lambda_1 + \lambda_2 > 1$ the optimal value of a_2 is 1 which means that in this range the objective function is strictly increasing in both multipliers. On the other hand, if $\lambda_1 + \lambda_2 < 1$, then the optimal value of a_2 is 0 implying an objective function that is strictly decreasing in λ_2. Hence, the optimal pairs of multipliers must sum to one. For any such pair the problem reduces to:

$$\inf_{\lambda_1\in[0,1]} \sup_{a_1\in[0,1]} -a_1 - c + \lambda_1(a_1 + c - b).$$

Since for any $\lambda_1 < 1$ the maximizing value of a_1 is 0 it follows that the objective is strictly decreasing in λ_1 (recall that $b > c$) and thus its unique optimal value is $\hat{\lambda}_1 = 1$. By our previous observations this implies $\hat{\lambda}_2 = 0$. Finally, notice that for this pair of multipliers every pair (a_1, a_2) solves the maximization in (24). Thus, $W^* = V^*$, but A^* is a strict subset of $A_{IS} = [0,1]^2$.

The recursive dual problem is:

$$W_1 = \inf_{\lambda_1\in\mathbb{R}_+} \sup_{a_1\in[0,1]} -a_1 + \lambda_1(a_1 - b) + W_2(\lambda_1), \qquad \text{where} \tag{25}$$

$$W_2(\lambda_1) = \inf_{\lambda_2\in\mathbb{R}_+} \sup_{a_2\in[0,1]} -a_2 + \lambda_1 a_2 + \lambda_2(a_2 - c).$$

We now show that the solutions from this problem coincide with A_{IS}. In the second period problem the optimal choice for the multiplier is $\hat{\lambda}_2(\lambda_1) = \max\{0, 1 - \lambda_1\}$. But then any $a_2 \in [0,1]$ is a maximizer and

$$W_2(\lambda_1) = \begin{cases} -(1-\lambda_1)c & \text{if } \lambda_1 < 1 \\ -(1-\lambda_1) & \text{otherwise.} \end{cases}$$

In the first period problem, if $\lambda_1 \in [0,1)$, then a_1 is optimally zero and the slope of the objective function with respect to λ_1 is $c - b < 0$. If $\lambda_1 > 1$, then a_1 is optimally 1 and the slope is $2 - b > 0$. Consequently, the unique optimal value for the multiplier is $\hat{\lambda}_1 = 1$. For this multiplier value the objective function is constant in a_1 and equal to $-b$.

Combining the results for the first and second period we thus find that the recursive method generates the same pair of Lagrange multipliers and the same solution set for the controls, namely $A_{IS} = [0,1]^2$. □

A sufficiency result: As Example 2 shows, the recursive dual policy correspondences may introduce extraneous solutions (i.e. they may have selections whose primal components do not maximize the sequential primal problem). On the other hand, if a saddle exists, there is a selection from these correspondences whose primal component solves the primal problem. Consequently, if there is a saddle and the recursive dual correspondences admit a unique primal selection, then this selection is a primal maximizer. Proposition 10 assumes a solution to the primal problem, the Slater condition and strictly concave objective and constraint functions. Primal existence, the Slater condition and concavity of the functions are classical conditions for the existence of a saddle point; strict concavity ensures a unique primal component to the recursive dual solution. Thus, the proposition guarantees that the recursive dual problem gives necessary and sufficient conditions for a sequential primal solution.

Remark 4. As discussed in Messner and Pavoni (2004), the recursive saddle point method of Marcet and Marimon (1999) fails to give sufficient conditions for a primal optimum. Marcet and Marimon (2011) also observe that this drawback can be overcome by imposing strict concavity assumptions on the functions that define the objective and the constraints. Cole and Kubler (2010) use dual state variables in a recursive formulation of the primal problem (they solve "sup-inf" not "inf-sup" problems at each step). To resolve sufficiency concerns in this context, Cole and Kubler (2010) augment the basic formulation with an expanded state space that includes end of period lottery realizations.

Assumption 1. *The sets A_1 and A_2 are convex and the functions f_1, g_1,m,l f_2, and g_2 are* *concave. Moreover, there exists an $\bar{a} \in A$ such that*

$$\begin{aligned} g_1(\bar{a}_1) + c_1 \cdot m(\bar{a}_2) &> 0 \\ c_2 \cdot l(\bar{a}_1) + g_2(\bar{a}_2) &> 0. \end{aligned}$$

Proposition 10. *(Sufficiency) Suppose that Assumption 1 is satisfied and that A^* is non-empty. Then: (i) The Lagrangian admits a saddle point and the value of the recursive dual problem coincides with the value of the optimization problem, i.e. $W_1 = V^*$. (ii) If, in addition, f_1 and f_2 are strictly concave, then A^* contains only one element and:*

$$\left[(\hat{\lambda}_1, \hat{\varphi}, \hat{\mu}, \hat{a}_1) \in \hat{\Gamma}_1 \quad \textit{and} \quad (\hat{\lambda}_2, \hat{a}_2) \in \hat{\Gamma}_2(\hat{\varphi}, \hat{\mu})\right] \Rightarrow \{(\hat{a}_1, \hat{a}_2)\} = A^*.$$

.

Proof. See Appendix. □

Remark 5. For $\alpha \in \mathbb{R}$, the α-upper (resp. lower) level set of a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by $\{x : h(x) \geq \alpha\}$ (resp. $\{x : h(x) \leq \alpha\}$). Primal attainment (the non-emptiness of $A^* = A_{SI}$) is ensured if the objective function f and constraint function g are upper semi-continuous and one or the other has a non-empty compact α-upper level set. More generally, the compact level set requirement on the functions f and g can be replaced by the requirement that for fixed $\bar{\lambda}$, the sequential Lagrangian $\mathscr{L}(\cdot, \bar{\lambda})$ has a non-empty compact α-upper level set in a. The Slater condition ensures that for the fixed value $\bar{a}$, the sequential Lagrangian $\mathscr{L}(\bar{a}, \cdot)$ has compact lower level sets. Together with the concavity assumptions and the finiteness of the optimal primal value, this condition guarantees that the value function P defined for this problem as in Subsection 2.1 has a sub-differential (and is convex and lower semicontinuous) at 0. Thus, by Proposition 4, dual attainment (the non-emptiness of Λ_{SI}) is ensured. Saddle point existence then follows from Proposition 1.

3.4 Duality Between Recursive Problems

In this section we derive additional duality relationships between primal and dual states and value functions. Since in our simple two-period setting states and value functions are defined only for the second period, we compare values and states only in this period. In the next section, we embed the problem into a framework that incorporates initial and terminal conditions. This allows us to define and derive relationships between first as well as second period value functions.

Recall the definitions of the second period value functions $V_2 : X \times Y \to \overline{\mathbb{R}}$ and $W_2 : M \times F \to \overline{\mathbb{R}}$ from the recursive primal and dual problems:

$$\begin{aligned} V_2(x,y) := \sup_{A_2} \inf_{\Lambda_2} \; & f_2(a_2) + \lambda_2 \cdot (c_2 \cdot x + g_2(a_2)) \\ \text{s.t.} \;\; & y = m(a_2); \end{aligned} \tag{26}$$

and

$$\begin{aligned} W_2(\mu,\phi) := \inf_{\Lambda_2} \sup_{A_2} \; & f_2(a_2) + \mu \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) \\ \text{s.t.} \;\; & \phi = \lambda_2 \cdot c_2. \end{aligned} \tag{27}$$

To relate these two problems we absorb their constraints $y = m(a_2)$ and $\phi = \lambda_2 \cdot c_2$ into Lagrangians and assume that these Lagrangians admit saddles. Formally, let $\gamma \in \mathbb{R}^{r_1}$ be the multiplier for the constraint $y = m(a_2)$ in (26) and $\zeta \in \mathbb{R}^{r_2}$ the multiplier for the constraint $\phi = \lambda_2 \cdot c_2$ in (27).

Definition 2 (Second period saddle points)**.** We say that the primal problem (26) admits a saddle at (x,y) if there exists a triple $(a_2^*; \lambda_2^*, \gamma^*) \in A_2 \times \Lambda_2 \times \mathbb{R}^{r_1}$ such that for all $(a_2; \lambda_2, \gamma) \in A_2 \times \Lambda_2 \times \mathbb{R}^{r_1}$:

$$\begin{aligned} & -\gamma^* \cdot y + f_2(a_2) + \lambda_2^* \cdot c_2 \cdot x + \lambda_2^* \cdot g_2(a_2) + \gamma^* \cdot m(a_2) \\ & \quad \le -\gamma^* \cdot y + f_2(a_2^*) + \lambda_2^* \cdot c_2 \cdot x + \lambda_2^* \cdot g_2(a_2^*) + \gamma^* \cdot m(a_2^*) \\ & \quad\quad \le -\gamma \cdot y + f_2(a_2^*) + \lambda_2 \cdot c_2 \cdot x + \lambda_2 \cdot g_2(a_2^*) + \gamma \cdot m(a_2^*). \end{aligned} \tag{28}$$

We say that the dual problem (27) admits a saddle at (ϕ, μ) if there exists a triple $(\lambda_2^*; a_2^*, \zeta^*) \in \Lambda_2 \times A_2 \times \mathbb{R}^{r_2}$ such that for all $(\lambda_2, a_2, \zeta) \in \Lambda_2 \times A_2 \times \mathbb{R}^{r_2}$:

$$\begin{aligned} & -\phi \cdot \zeta + f_2(a_2) + \mu \cdot m(a_2) + \lambda_2^* \cdot g_2(a_2) + \lambda_2^* \cdot c_2 \cdot \zeta \\ & \quad \le -\phi \cdot \zeta^* + f_2(a_2^*) + \mu \cdot m(a_2^*) + \lambda_2^* \cdot g_2(a_2^*) + \lambda_2^* \cdot c_2 \cdot \zeta^* \\ & \quad\quad \le -\phi \cdot \zeta^* + f_2(a_2^*) + \mu \cdot m(a_2^*) + \lambda_2 \cdot g_2(a_2^*) + \lambda_2 \cdot c_2 \cdot \zeta^*. \end{aligned} \tag{29}$$

The next result uses conjugate duality to relate the value functions from the continuation problems: modulo sign changes, the conjugate of one is a lower bound for the other. The value function of one equals the conjugate of the other whenever the first problem admits a saddle point.

Proposition 11. *(i) For all $(x, y) \in X \times Y$ and $(\mu, \phi) \in M \times F$ the following inequalities hold:*

$$W_2(\mu, \phi) \geq (-V_2)^\star(\mu, -\phi) \quad (30)$$

$$-V_2(x, y) \geq (W_2)^\star(-x, y). \quad (31)$$

(ii) the inequality in (30) (resp. (31)) holds with equality whenever the dual second period problem (27) admits a saddle at (μ, ϕ) (resp. whenever the primal second period problem (26) admits a saddle at (x, y)).

Proof. See Appendix. □

The next proposition shows that if the underlying sequential optimization problem admits a saddle, then both the primal (26) and dual (27) continuation problems admit saddles *at the states induced by the sequential saddle*.

Proposition 12. *Assume that the underlying sequential optimization problem admits a saddle (a^*, λ^*) and let $(\mu^*, \phi^*) = (\lambda_1^* \cdot c_1, \lambda_2^* \cdot c_2)$ and $(y^*, x^*) = (m(a_2^*), l(a_1^*))$. Then: (i) the dual continuation problem (27) admits a saddle at (μ^*, ϕ^*), and the primal continuation problem (26) admits a saddle at (x^*, y^*); (ii) W_2 (resp. $-V_2$) is sub-differentiable at (μ^*, ϕ^*) (resp. (x^*, y^*)) and:*

$$\begin{bmatrix} -\phi^* \\ \mu^* \end{bmatrix} \in \partial\left[-V_2\left(x^*, y^*\right)\right] \quad \text{and} \quad \begin{bmatrix} y^* \\ -x^* \end{bmatrix} \in \partial W_2\left(\mu^*, \phi^*\right).$$

Proof. See Appendix. □

The second part of Proposition 12 implies that, modulo sign changes, optimal second period dual states (μ^*, ϕ^*) are sub-differentials of the second period primal value function at the optimal primal states (x^*, y^*) and vice versa.[14] In applications, this relationship allows an economic interpretation to be placed upon optimal dual states: they are shadow prices of optimal primal states.

Further remarks on state spaces Note that the second period primal problem is well posed and has a non-empty constraint set if and only if (x, y) belongs to:

$$S := \{(x, y) \in X \times Y | \exists a_2 \in A_2, y = m(a_2), c_2 \cdot x + g_2(a_2) \geq 0\} \subset X \times Y \subset \mathbb{R}^{r_1 + r_2}.$$

Outside of S, V_2 equals $-\infty$. From the perspective of practical computation, non-finite valued functions are awkward and it is natural to treat S as the (endogenous) second

[14] In the proof we also show that the second period sequential saddle variables, (a_2^*, λ_2^*) are saddles of the second period problems.

period state space. S may, however, be difficult to characterize. In particular, this is often the case when m and g_2 are bounded functions ensuring that S is a more restricted set.

The dual analogue of S is:

$$R := \{(\mu, \phi) \in M \times F | W_2(\mu, \phi) < \infty\}.$$

If f_2 and g_2 are bounded above and m is bounded, then $R = M \times F$. Since M and F are the images of the positive cones $\mathbb{R}^{k_1}_+$ and $\mathbb{R}^{k_2}_+$ under the linear maps c_1 and c_2, in this case R is easy to determine. This is a significant practical advantage of the recursive dual approach.

As a final remark, we note that Proposition 12 indicates an interesting duality in the 'direction' (forward or backward) of states. Optimal *forward* primal states are subdifferentials of optimal *backward* dual states and vice versa.

4 Initial and Terminal Conditions for States

In this section, we embed our simple two-period problem into a family of problems parameterized by initial and terminal conditions for state variables. This extension clarifies how the recursive dual approach is applied in general multi-period settings. It also makes transparent an issue concealed by our simple model. Since forward-looking state variables capture the relevant aspects of continuation paths of controls or multipliers in a sequential problem, they may have a natural terminal condition, but they do not have a natural initial condition. Thus, the initial values of forward-looking states are determined endogenously. We show how this is done and what it implies for the recursive approaches.

The Primal Problem. We augment the constraints of the sequential primal problem (9) with parameters $x_1 \in \mathbb{R}^{r_1}$ and $y_3 \in \mathbb{R}^{r_2}$ to give the family of perturbed problems:

$$\begin{aligned} P(x_1, y_3) = \sup_{a \in A} \quad & f_1(a_1) + f_2(a_2) \\ \text{s.t.} \quad & g_1(a_1) + c_1 \cdot (m(a_2) + y_3) \geq 0 \\ & c_2 \cdot (l(a_1) + x_1) + g_2(a_2) \geq 0. \end{aligned} \tag{32}$$

In addition, to save on notation and without loss of generality, we assume that for $t = 1, 2$, $A_t = A \subset \mathbb{R}^n$ and $g_t : A \to Y = \mathbb{R}^k$. The parameter x_1 belongs to the range of the function l and has a natural interpretation as an initial value for the backwards-looking primal state variable. Similarly, y_3 belongs to the range of the function m and has a natural interpretation as a terminal value for the forwards-looking primal state variable.

In what follows we focus on problems in which $y_3 = 0$. However, for the purposes of symmetry and generality, we explicitly include y_3 in the notation. Clearly, by setting x_1 as well as y_3 to 0, we recover our original primal problem (9). We explicitly define backwards and forwards-looking primal state variables for the other dates by writing, for $t = 1, 2$,

$$x_{t+1} = x_t + l(a_t) \text{ and } y_t = m(a_t) + y_{t+1}$$

with x_1 and y_3 given. Thus, $x_t = x_1 + \sum_{\tau=1}^{t} l(a_\tau)$, $t = 2, 3$, gives the accumulation of $l(a_\tau)$ terms at t starting from x_1 and $y_t = \sum_{\tau=t}^{2} m(a_\tau) + y_3$, $t = 1, 2$, gives the accumulation of $m(a_\tau)$ terms at t starting from y_3.

The primal recursive approach to solving (32) involves two steps.[15] In the first a pair, $t = 1, 2$, (more generally a sequence) of recursive problems is solved:

$$V_t(x_t, y_t) = \sup_{A\times X\times Y} \inf_{\Lambda} f_t(a_t) + \lambda_t \cdot (b_{1,t}x_t + g_t(a_t) + b_{2,t} \cdot y_{t+1}) + V_{t+1}(x_{t+1}, y_{t+1}) \tag{33}$$

$$\text{s.t. } x_{t+1} = x_t + l(a_t), \quad y_t = m(a_t) + y_{t+1} \text{ and } x_t, y_t \text{ given,}$$

where $b_{1,1} = 0$, $b_{1,2} = c_2$, $b_{2,1} = c_1$, $b_{2,2} = 0$, $V_3(x_3, y_3) = 0$ if $y_3 = 0$ and $V_3(x_3, y_3) = -\infty$ otherwise and $\Lambda = \mathbb{R}^k_+$. In the second step, the optimal value of y_1 is found:

$$P(x_1, 0) = \sup_{Y} V_1(x_1, y_1).$$

In particular, the optimal value from the unperturbed problem considered in earlier sections is given by $V^* = P(0, 0) = \sup_Y V_1(0, y)$. It is clear that this procedure can be extended to any finite horizon choice problem with initial and terminal conditions, a fact which is proved using an elaboration of the argument underpinning Proposition 6.[16, 17]

[15] This may be formally shown along the lines of Proposition 6.

[16] In our current formulation constraints involving only forwards-looking state variables are applied in the first period, while constraints involving only backwards-looking state variables are applied in the second. More generally, both types of constraint or constraints involving both types of state variable may be applied in the same period. A small extension of our formulation handles such problems.

[17] For a treatment of the infinite horizon problem, see Messner, Pavoni, and Sleet (2011).

The Dual Problem. We now consider a *dual problem* with dual initial and terminal conditions. First, recall the dual problem without such conditions:

$$W^* = \inf_{\Lambda^2} \sup_{A} \quad f_1(a_1) + f_2(a_2) + \lambda_1 \left[g_1(a_1) + c_1 \cdot m(a_2) \right] \\ + \lambda_2 \left[c_2 \cdot l(a_1) + g_2(a_2) \right]. \tag{34}$$

The recursive version of this problem used the state variables (and laws of motion for state variables) $\mu_2 = \lambda_1 \cdot c_1$ and $\phi_2 = \lambda_2 \cdot c_2$. Inserting this notation into (34) gives:

$$W^* = \inf_{\Lambda^2 \times M \times F} \sup_{A} \quad f_1(a_1) + f_2(a_2) + \lambda_1 \cdot g_1(a_1) + \mu_2 \cdot m(a_2) \\ + \phi_2 \cdot l(a_1) + \lambda_2 \cdot g_2(a_2) \tag{35}$$

subject to: $\mu_2 = \lambda_1 \cdot c_1$ and $\phi_2 = \lambda_2 \cdot c_2$. The variables μ_2 and ϕ_2 can be interpreted as the shadow prices of $y_2 = m(a_2) + y_3$ and $x_2 = x_1 + l(a_1)$ in a setting in which y_3 and x_1 equal 0. The dual analogue of (32) augments (35) with dual state variables μ_1 and ϕ_3. These can be viewed as initial backward-looking and terminal forward-looking dual states, respectively, or, equivalently, shadow prices for $y_1 = m(a_1) + m(a_2) + y_3$ and $x_3 = x_1 + l(a_1) + l(a_2)$ where, once again, y_3 and x_1 are set to zero. The dual analogue of (32) is:

$$Q(\mu_1, \phi_3) = \inf_{\Lambda^2 \times M \times F} \sup_{A} \quad f_1(a_1) + f_2(a_2) + \lambda_1 \cdot g_1(a_1) + \mu_1 m(a_1) + \mu_2 \cdot m(a_2) \\ + \phi_2 \cdot l(a_1) + \lambda_2 \cdot g_2(a_2) + \phi_3 \cdot l(a_2) \tag{36}$$

subject to: $\mu_2 = \mu_1 + \lambda_1 \cdot c_1$ and $\phi_2 = \lambda_2 \cdot c_2 + \phi_3$. Of course, $W^* = Q(0, 0)$.

A two step procedure similar to the primal case may be used to solve for $Q(\cdot, 0)$ and, hence, W^*. First the recursive dual problems, for $t = 1, 2$, are solved:

$$W_t(\mu_t, \phi_t) = \inf_{\Lambda \times M \times F} \sup_{A} f_t(a_t) + \mu_t \cdot m(a_t) + \lambda_t \cdot g_t(a_t) + \phi_{t+1} \cdot l(a_t) + W_{t+1}(\mu_{t+1}, \phi_{t+1}) \\ \text{s.t. } \phi_t = \lambda_t \cdot b_{1,t} + \phi_{t+1}, \quad \mu_{t+1} = \mu_t + \lambda_t \cdot b_{2,t} \text{ and } \mu_t, \phi_t \quad \text{given}, \tag{37}$$

where $W_3(\mu_3, \phi_3) = 0$ if $\phi_3 = 0$ and $W_3(\mu_3, \phi_3) = \infty$ otherwise. Second, the optimal value of ϕ_1 is found:

$$Q(\mu_1, 0) = \inf_{F} W_1(\mu_1, \phi_1), \tag{38}$$

where the equality (38) may be formally proved using an elaboration of the argument underpinning Proposition 7. In particular, the optimal value from the unperturbed dual problem (34) is given by: $W^* = \inf_F W_1(0, \phi)$. As in the primal case, this two step procedure can be extended to any finite horizon problem. In addition, along similar lines to Proposition 8, it may be shown that solutions so obtained solve (34). If the underlying sequential problem admits a saddle point, then this procedure also yields an optimal primal value and necessary conditions for optimal primal solutions.

We conclude this section with a conjugacy result for the first period value functions V_1 and W_1. It is analogous to the result for V_2 and W_2 stated in Proposition 11. As a precursor we define Lagrangians $\mathscr{L}^P$ and $\mathscr{L}^D$ for the sequential counterparts of the recursive primal (33) and dual (37) problems:[18]

$$\mathscr{L}^P(a, \lambda, \gamma; \bar{x}_1, y_1) = H(a, \lambda) + \left(\sum_{\tau=1}^{2} \lambda_\tau b_{1,\tau} \right) \cdot \bar{x}_1 + \gamma \cdot \left[\sum_{\tau=1}^{2} m(a_\tau) - y_1 \right], \tag{39}$$

$$\mathscr{L}^D(a, \lambda, z; \bar{\mu}_1, \phi_1) = H(a, \lambda) + \bar{\mu}_1 \cdot \left(\sum_{\tau=1}^{2} m(a_\tau) \right) + z \cdot \left[\sum_{\tau=1}^{2} \lambda_\tau \cdot b_{1,\tau} - \phi_1 \right], \tag{40}$$

where

$$H(a, \lambda) = f_1(a_1) + f_2(a_2) + \lambda_1 \cdot [g_1(a_1) + b_{2,1} \cdot m(a_2)] + \lambda_2 \cdot [b_{1,2} \cdot l(a_1) + g_2(a_2)].$$

Since the multipliers γ and z are associated with the equality constraints $y_1 = m(a_1) + m(a_2)$ and $\phi_1 = \sum_{\tau=1}^{2} \lambda_\tau \cdot b_{1,\tau}$ they are not restricted to be positive.

Proposition 13. *Let $\bar{x}_1$ and $\bar{\mu}_1$ be given,*

$$\hat{V}^*(\bar{\mu}_1, \bar{x}_1) := \sup_{Y} \quad V_1(\bar{x}_1, y_1) + \bar{\mu}_1 \cdot y_1, \tag{41}$$

$$\hat{W}^*(\bar{\mu}_1, \bar{x}_1) := \inf_{F} \quad W_1(\bar{\mu}_1, \phi_1) + \phi_1 \cdot \bar{x}_1, \tag{42}$$

y_1^* solve (41) and ϕ_1^* solve (42). Moreover, let a^* and λ^* be solutions of, respectively,

$$\max_{A} \inf_{\Lambda, \Gamma} \mathscr{L}^P(a, \lambda, \gamma; \bar{x}_1, y_1^*) \qquad \textit{and} \qquad \min_{\Lambda} \sup_{A, Z} \mathscr{L}^D(a, \lambda, z; \bar{\mu}_1, \phi_1^*).$$

(i) $\hat{V}^*(\bar{\mu}_1, \bar{x}_1) \leq \hat{W}^*(\bar{\mu}_1, \bar{x}_1)$.
(ii) $\hat{V}^*(\bar{\mu}_1, \bar{x}_1) = \hat{W}^*(\bar{\mu}_1, \bar{x}_1)$ if and only if $(a, \lambda, \gamma) = (a^*, \lambda^*, \bar{\mu}_1)$ is a saddle of $\mathscr{L}^P(a, \lambda, \gamma; \bar{x}_1, y_1^*)$

[18] Using our insights from the preceding sections it is straightforward to show that for all (x_1, y_1) we have $V_1(x_1, y_1) = \sup_A \inf_{\Lambda,\Gamma} \mathscr{L}^P(a, \lambda, \gamma; x_1, y_1)$; similarly, for all (μ_1, ϕ_1), $W_1(\mu_1, \phi_1) = \inf_\Lambda \sup_{A,Z} \mathscr{L}^D(a, \lambda, z; \mu_1, \phi_1)$

and $(a, \lambda, z) = (a^*, \lambda^*, \bar{x}_1)$ *is a saddle of* $\mathscr{L}^D(a, \lambda, z; \bar{\mu}_1, \phi_1^*)$.
(iii) If $\hat{V}^*(\bar{\mu}_1, \bar{x}_1) = \hat{W}^*(\bar{\mu}_1, \bar{x}_1)$, *then* $W_1(\bar{\mu}_1, \phi_1^*) = (-V_1)^\star(\bar{\mu}_1, \phi_1^*)$, *and* $-V_1(\bar{x}_1, y_1^*) = (W_1)^\star(-\bar{x}_1, y_1^*)$. *Moreover, under the same conditions,* W_1 *and* V_1 *are sub-differentiable at* $(\bar{\mu}_1, \phi_1^*)$ *and* $(\bar{x}_1, y_1^*)$, *respectively, with:*

$$\begin{bmatrix} -\phi_1^* \\ \bar{\mu}_1 \end{bmatrix} \in \partial \left[-V_1\left(\bar{x}_1, y_1^*\right)\right] \quad \text{and} \quad \begin{bmatrix} y_1^* \\ -\bar{x}_1 \end{bmatrix} \in \partial W_1\left(\bar{\mu}_1, \phi_1^*\right).$$

Proof. See Appendix. □

Evidently, $V^* = \hat{V}^*(0,0)$ and $W^* = \hat{W}^*(0,0)$. More generally, Proposition 13 shows how initial values for the forward state variables y_1 or ϕ_1 may be determined given arbitrary initial conditions for the backward states $(\bar{\mu}_1, \bar{x}_1)$. In addition, Proposition 13 combined with Proposition 12 implies that dual relationships hold along the saddle path whenever a saddle exists for the primal problem with initial condition $\bar{x}_1$.

5 Additive applications

We have developed the dual recursive method in a two period setting. It is easily extended to multiple, but finite numbers of periods. We now give two multi-period economic applications.

5.1 An AK-growth model

Our first example is a standard, finite horizon AK-growth model. A T-period lived agent has preferences over a consumption stream $\{c_t\}_{t=1}^T \in \mathbb{R}_+^T$ given by $\sum_{t=1}^T u_t(c_t)$. She allocates current output to consumption c_t and investment i_t; she has a linear technology with parameter $A > 0$ for converting capital (accumulated undepreciated past investment) into output. We place her choice problem directly into the form (9), writing it as:

$$V^* = \sup_{\{c_t, i_t\}_{t=1}^T \in \mathbb{R}_+^{2T}} \sum_{t=1}^T u_t(c_t) \tag{43}$$

$$\text{s.t.} \quad x_1 = \bar{x}_1 \quad \text{and} \quad t = 1, \cdots, T, \quad c_t + i_t \leq A\left((1-\delta)^{t-1} x_1 + \sum_{\tau=1}^{t-1} (1-\delta)^{\tau-1} i_\tau\right),$$

where $\delta \in [0,1)$ is depreciation. To economize on notation, assume that $\delta = 0$ and $A = 1$. Problem (43) may then be re-expressed in sup-inf form as:

$$V^* = \sup_{\mathbb{R}_+^{2T}} \inf_{\Lambda} \sum_{t=1}^{T} u_t(c_t) + \lambda_t \left[\left(\bar{x}_1 + \sum_{\tau=1}^{t-1} i_\tau \right) - c_t - i_t \right], \tag{44}$$

where $\Lambda := \mathbb{R}_+^T$. Defining capital as the sum of past investments, $x_t = x_1 + \sum_{\tau=1}^{t-1} i_\tau$, the constraints can be re-expressed as, for all t, $c_t + i_t \leq x_t$, $x_{t+1} = x_t + i_t$ with x_1 given. This is the form they are usually given in (i.e. in most growth problems, the state variable capital is given explicitly as a primitive of the problem). In our nomenclature, capital is a backwards primal state variable. The associated recursive primal formulation then generates the sequence of functional equations:

$$V_t(x_t) = \sup_{\mathbb{R}_+^3} \inf_{\mathbb{R}_+} u_t(c_t) + \lambda_t(x_t - c_t - i_t) + V_{t+1}(x_{t+1})$$
$$\text{s.t. } x_{t+1} = x_t + i_t, \quad \text{and} \quad x_t \quad \text{given,}$$

with $V_{T+1} \equiv 0$. The optimal value of the original problem, V^*, is obtained by evaluating V_1 at the initial value of the capital stock, $x_1 = \bar{x}_1$. That is, $V_1(\bar{x}_1) = V^*$.

The associated dual problem is:

$$W^* = \inf_{\Lambda} \sup_{\mathbb{R}_+^{2T}} \sum_{t=1}^{T} u_t(c_t) + \lambda_t \left[\left(\bar{x}_1 + \sum_{\tau=1}^{t-1} i_\tau \right) - c_t - i_t \right]. \tag{45}$$

Defining $\phi_t = \sum_{\tau=t}^{T} \lambda_\tau$, so that $\phi_t = \phi_{t+1} + \lambda_t$ with $\phi_{T+1} = 0$, problem (45) can be re-expressed as:

$$W^* = \inf_{\mathbb{R}_+^{2T}} \sup_{\mathbb{R}_+^{2T}} \sum_{t=1}^{T} [u_t(c_t) + \phi_{t+1} i_t - \lambda_t(c_t + i_t)] + \phi_1 \bar{x}_1 \tag{46}$$
$$\text{s.t. } t = 1, \cdots, T, \quad \phi_t = \lambda_t + \phi_{t+1}, \quad \phi_{T+1} = 0.$$

Notice that the variables ϕ_t are forward-looking summaries of future multipliers. Problem (46) may be solved in two steps. In the first, it is recursively decomposed and the following sequence of problems, $t = 1, \cdots, T$, is solved:

$$W_t(\phi_t) = \inf_{(\lambda_t, \phi_{t+1}) \in \mathbb{R}_+^2} \sup_{(c_t, i_t) \in \mathbb{R}_+^2} u_t(c_t) + \phi_{t+1} i_t - \lambda_t (c_t + i_t) + W_{t+1}(\phi_{t+1}) \tag{47}$$
$$\text{s.t. } \phi_t = \lambda_t + \phi_{t+1},$$

with $W_{T+1}(\phi) = 0$ if $\phi = 0$ and $W_{T+1}(\phi) = \infty$ otherwise. Notice that in passing from Problem (46) to Problem (47), the term $\phi_1 \bar{x}_1$, is omitted.

Since, the forward-looking dual state variables $\{\phi_t\}$ are not subject to an initial condition, the optimal ϕ_1 value is solved for in a second step by calculating:

$$\inf_{\phi_1 \in \mathbb{R}_+} W_1(\phi_1) + \phi_1 \bar{x}_1.$$

5.2 Limited Commitment

Our primal formulation of the AK-model featured a backward-looking state variable and our dual formulation a forward-looking one. We now consider a limited commitment model in which these directions are reversed.

Let $\{a_t\}_{t=1}^T \in A^T$ denote a sequence of actions taken by an agent, let $u_t : A \to D \subset \mathbb{R}$ denote the agent's period t utility function and $r_t : A \to \mathbb{R}$ a principal's period t return function. Let b_t denote a period t outside utility option for the agent. The principal's problem is:

$$V^* = \max_{A^T} \sum_{t=1}^{T} r_t(a_t) \tag{48}$$
$$\text{s.t. } \sum_{\tau=t}^{T} u_\tau(a_\tau) \geq b_t \qquad t = 1, \ldots T.$$

This problem can be written in sup-inf form as:

$$V^* = \sup_{A^T} \inf_{\Lambda} \sum_{t=1}^{T} \left[r_t(a_t) + \lambda_t \left(\sum_{\tau=t}^{T} u_\tau(a_\tau) - b_t \right) \right], \tag{49}$$

where $\Lambda = \mathbb{R}_+^T$. Defining y_t to be the utility attached by the agent to a future action stream, i.e. $y_t = \sum_{\tau=t}^T u_\tau(a_\tau)$ for some $\{a_\tau\}_{\tau=t}^T$, the constraints in (48) can be re-expressed as, for $t \leq T$, $u(a_t) + y_{t+1} \geq b_t$ with $y_{T+1} = 0$ given. In our terminology y_t is a forwards-looking primal state variable (with a terminal condition).

Problem (49) can be solved using the two step procedure applied to the dual AK problem. First, (49) is recursively decomposed using the value functions $V_t : \mathbb{R} \to \overline{\mathbb{R}}$ as:

$$V_t(y_t) = \sup_{A \times \mathbb{R}} \inf_{\mathbb{R}_+} r_t(a_t) + \lambda_t(u_t(a_t) + y_{t+1} - b_t) + V_{t+1}(y_{t+1}) \tag{50}$$
$$\text{s.t.} \quad u_t(a_t) + y_{t+1} = y_t,$$

where $V_{T+1}(y) = 0$ if $y = 0$ and $V_{T+1}(y) = -\infty$ otherwise. Notice that at each date t there is a set of utilities Y_t such that V_t is finite-valued. Specifically, $V_t(y_t)$ is finite if and only if $y_t \geq b_t$ and y_t is in the range of $\sum_{\tau=t}^{T} u_\tau(a_\tau)$. Y_t represents the effective (endogenous) state space of the t-th period problem. Since non-finite value functions are problematic from a practical computational point of view, it is useful to jointly compute V_t and Y_t when solving (50). This is not difficult in the simple current setting, but often is in more elaborate settings.

Solving the problems (50) gives the period 1 value function, V_1. Since there is no parameter y_1 constraining the principal's choice in (48), the optimal value of y_1 must be solved for in a second step. This is done by calculating:

$$\max_{y_1 \in \mathbb{R}} V_1(y_1). \tag{51}$$

The dual problem associated with (49) is:

$$W^* = \inf_{\Lambda} \sup_{A^T} \sum_{t=1}^{T} \left[r_t(a_t) + \lambda_t \left(\sum_{\tau=t}^{T} u_\tau(a_\tau) - b_t \right) \right]. \tag{52}$$

Problem (52) can be formulated recursively using summaries of past Lagrange multipliers as state variables. Unlike the primal state variable (continuation utility), these dual states are backward looking. Defining $\mu_t = \sum_{\tau=1}^{t-1} \lambda_\tau$, problem (52) can be re-expressed as:

$$W^* = \inf_{\mathbb{R}_+^{2T}} \sup_{A^T} \sum_{t=1}^{T} [r_t(a_t) + \lambda_t(u_t(a_t) - b_t) + \mu_t u_t(a_t)] \tag{53}$$
$$\text{s.t. } \mu_1 = 0 \ \text{ and } \ t = 1, \cdots, T, \ \mu_{t+1} = \mu_t + \lambda_t.$$

Problem (53) can be recursively decomposed as:

$$W_t(\mu_t) = \inf_{\mathbb{R}^2_+} \sup_{A} r_t(a_t) + \mu_t u_t(a_t) + \lambda_t(u_t(a_t) - b_t) + W_{t+1}(\mu_{t+1}) \tag{54}$$
$$\text{s.t. } \mu_{t+1} = \mu_t + \lambda_t, \quad \mu_t \quad \text{given,}$$

with $W_{T+1} \equiv 0$. The optimal value W^* is obtained by evaluating W_1 at the initial condition $\mu_1 = 0$.

6 Non-additively separability: Dynamic Moral Hazard

So far we have restricted attention to problems with an additively separable structure. Although this encompasses a large set of economic applications, it omits many others. In particular, it omits many dynamic moral hazard problems that have a recursive primal formulation. In this section we use a standard dynamic moral hazard model to illustrate the additional issues that arise in the application of the recursive dual approach to non-additively separable problems. In contrast to the additively separable case, the interchange of conditional inf and sup operations that is needed to pass from the sequential to the recursive dual is no longer guaranteed to preserve the optimal dual value or the optimal sequential dual solution set. Values are preserved if a sequential saddle exists in which case the optimal recursive dual value is also the optimal sequential primal value.

6.1 A simple moral hazard problem

In each of two periods an agent chooses an effort level from a finite set $E = \{e_k\}_{k=1}^K$. The effort of the agent induces a probability distribution over a set of outputs with N elements. The principal offers a contract that specifies the agent's effort and wage in each period. Formally, a contract is a tuple $\{e_1, \{w_1^i\}, \{e_2^i\}, \{w_2^{ij}\}\}$, where $e_1 \in E$ is first period effort, $w_1^i \in \mathbb{R}_+$ is the first period wage following the i-th first period output realization, $e_2^i \in E$ is the second period effort following this output realization and $w_2^{ij} \in \mathbb{R}^N_+$ is the second period wage given the i-th first and j-th second period output realizations. Let $A_1 = \widetilde{A} = E \times \mathbb{R}^N_+$ denote the set of first period contracts, $A_2 = \widetilde{A}^N$ the set of second period contracts and $A = A_1 \times A_2$ the set of lifetime contracts. Elements of these sets are denoted a_1, $a_2 = \{a_2^i\}$ and a respectively. The probability distribution over outcomes induced by effort e is denoted $\pi(e) = (\pi_1(e), \dots, \pi_N(e))$. $r^i(w)$ is the per period payoff of the principal when outcome i is realized and the wage payment w is made. If the agent exerts effort $e \in E$ and receives wage $w \in \mathbb{R}_+$ then his (per period) payoff is $u(w, e)$. If the agent does not accept the contract proposed by the principal he receives a payoff of $\bar{u}$.

To compare this moral hazard model with the optimization problem considered in the preceding sections, define the following functions for $a, a' \in \tilde{A}$ and $\hat{e} \in E$:

$$
\begin{aligned}
& f(a) := \sum_{i=1}^{N} \pi_i(e) r^i(w^i) \quad \text{and} \quad m(a) := \sum_{i=1}^{N} \pi_i(e) u(w^i, e), \\
& \bar{g}_1^i(a, m(a')) := u(w^i, e) + m(a') - \bar{u}, \\
& g_1^{\hat{e},i}(a, m(a')) := u(w^i, e) + m(a') - \frac{\pi_i(\hat{e})}{\pi_i(e)} \left[u(w^i, \hat{e}) + m(a') \right], \\
& g_2^{\hat{e}}(a) := \sum_{i=1}^{N} \pi_i(e) u(w^i, e) - \sum_{i=1}^{N} \pi_i(\hat{e}) u(w^i, \hat{e}).
\end{aligned}
$$

Using this notation the principal's primal optimization problem is:

$$
\begin{aligned}
V^* = \max_{a \in A} \ & f(a_1) + \sum_{i=1}^{N} \pi_i(e_1) f(a_2^i) \\
\text{s.t.} \ & \sum_{i=1}^{N} \pi_i(e_1) \bar{g}_1^i(a_1, m(a_2^i)) \geq 0 \\
& \sum_{i=1}^{N} \pi_i(e_1) g_1^{\hat{e},i}(a_1, m(a_2^i)) \geq 0 && \text{for all } \hat{e} \in E, \\
& g_2^{\hat{e}}(a_2^i) \geq 0, && \text{for all } \hat{e} \in E \text{ and all } i = 1, \ldots, N.
\end{aligned} \tag{55}
$$

The first constraint is an individual rationality condition; the second and third are incentive compatibility conditions for first and second period effort.

The key difference between the problems considered in previous sections and the moral hazard problem (55) is that in the latter neither the objective function nor the constraints are additively separable in a_1 and a_2. Since the contract offered by the principal conditions on the outcome history it follows that the continuation payoffs of the principal and of the agent depend on past effort levels multiplicatively via the distribution over past outcomes. The Lagrangian associated with (55) is:

$$
\mathscr{L}(a, \lambda) := f(a_1) + \sum_{i=1}^{N} \pi_i(e_1) \left[f(a_2^i) + \bar{\lambda}_1 \bar{g}_1^i(a_1, m(a_2^i)) + \sum_{\hat{e} \in E} \lambda_1^{\hat{e}} g_1^{\hat{e},i}(a_1, m(a_2^i)) + \sum_{\hat{e} \in E} \lambda_2^{\hat{e},i} g_2(a_2^i) \right]. \tag{56}
$$

The primal problem can be re-expressed as:

$$V^* := \sup_{A} \inf_{\Lambda} \mathscr{L}(a, \lambda)$$

and recursively decomposed using the agent's continuation utilities $y^i = m(a_2^i)$ as (forward-looking) states. Letting $Y = m(\tilde{A})$, $\Lambda_1 = \mathbb{R}_+^{K+1}$ and $\Lambda_2 = \mathbb{R}_+^K$, we obtain the recursive primal problem:

$$V_1 \quad := \sup_{A_1 \times Y^N} \inf_{\Lambda_1} \quad f(a_1) + \sum_i \pi_i(e_1) \left[V_2(y^i) + \bar{\lambda}_1 \bar{g}_1^i(a_1, y^i) + \sum_{\hat{e} \in E} \lambda_1^{\hat{e}} g_1^{\hat{e},i}(a_1, y^i) \right],$$

where:

$$V_2(y^i) := \sup_{\widetilde{A}} \inf_{\Lambda_2} \; f(a_2^i) + \sum_{\hat{e} \in E} \lambda_2^{\hat{e},i} g_2^{\hat{e}}(a_2^i) \qquad \text{s.t. } y^i = m(a_2^i). \tag{57}$$

A variation of the argument underpinning Proposition 6 establishes that $V_1 = V^*$ and that the set of solutions generated by the recursive primal problem coincides with the set of solutions from the sequential primal.

We next consider the problem of decomposing the dual problem:

$$W^* := \inf_{\Lambda} \sup_{A} \mathscr{L}(a, \lambda).$$

To do so it is convenient to rewrite the Lagrangian in the form:

$$\begin{aligned} \mathscr{L}(a, \lambda) = f(a_1) + \sum_{i=1}^{N} \pi_i(e_1) \Bigg\{ & \bar{\lambda}_1 (u(w_1^i, e_1) - \bar{u}) + \sum_{\hat{e}} \lambda_1^{\hat{e}} \left(u(w_1^i, e_1) - \frac{\pi_i(\hat{e})}{\pi_i(e_1)} u(w_1^i, \hat{e}) \right) \\ & + f(a_2^i) + \sum_j \pi_{j=1}^N(e_2^i) \left[\left(\bar{\lambda}_1 + \sum_{\hat{e}} \lambda_1^{\hat{e}} \frac{\pi_i(e_1) - \pi_i(\hat{e})}{\pi_i(e_1)} \right) u(w_2^{ij}, e_2^i) \right. \\ & \left. + \sum_{\hat{e}} \lambda_2^{\hat{e},i} \left(u(w_2^{ij}, e_2^i) - \frac{\pi_j(\hat{e})}{\pi_j(e_2^i)} u(w_2^{ij}, \hat{e}) \right) \right] \Bigg\} \end{aligned}$$

A 'natural' way to formulate this problem recursively is to use (backward looking) state variables μ^i to represent $\bar{\lambda}_1 + \sum_{\hat{e}} \lambda_1^{\hat{e}} (\pi_i(e_1) - \pi_i(\hat{e})) / \pi_i(e_1)$. The latter term carries relevant information about first period choices into the second period. Using the state variable $\mu = \{\mu^i\}$, we can write:

$$W_1 := \inf_{\Lambda_1} \sup_{A_1} f(a_1) + \sum_{i=1}^{N} \pi_i(e_1) \Big[\bar{\lambda}_1 (u(w^i, e_1) - \bar{u}) \\ + \sum_{\hat{e} \in E} \lambda_1^{\hat{e}} \left(u(w_1^i, e_1) - \frac{\pi_i(\hat{e})}{\pi_i(e_1)} u(w_1^i, \hat{e}) \right) + W_2(\mu^i(\lambda_1, e_1)) \Big], \tag{58}$$

where $\mu^i(\lambda_1, e_1) = \bar{\lambda}_1 + \sum_{\hat{e} \in E} \lambda_1^{\hat{e}} \frac{\pi_i(e_1) - \pi_i(\hat{e})}{\pi_i(e_1)}$ and:

$$W_2(\mu^i) := \inf_{\Lambda_2} \sup_{\widetilde{A}} f(a_2^i) + \sum_{j=1}^{N} \pi_j(e_2) \Big[\mu^i u_2 \left(w_2^{ij}, e_2 \right) \\ + \sum_{\hat{e} \in E} \lambda_2^{\hat{e},i} \left(u(w_2^{i,j}, e_2^i) - \frac{\pi_j(\hat{e})}{\pi_j(e_2^i)} u(w_2^{i,j}, \hat{e}) \right) \Big]. \tag{59}$$

Dual policy correspondences $\hat{\Gamma}_1$ and $\hat{\Gamma}_2$ are then defined analogously to before.

Remark 6. In the additively separable case, recursive dual state variables depend only upon multipliers. In the moral hazard model, and more generally in non-additive problems, recursive dual state variables depend on and summarize information about both dual and primal variables. Note that we continue to refer to (58) as a recursive dual problem because it involves sequences of inf-sup not sup-inf variables and because state variables perturb the objective not the constraints.

Recall that in establishing the equivalence of the sequential and recursive dual problems in the additive case, we interchanged a conditional inf and a sup operation (see Remark 2). While this was immediately possible in the additively separable case, it is not in the non-additive case.[19] It is possible if the Lagrangian admits a sequential saddle.

Proposition 14. *If the sequential Lagrangian (56) admits a saddle (a^*, λ^*), then i) $W_1 = W^*$ and ii) $(\lambda_1^*, a_1^*) \in \hat{\Gamma}_1$ and $(\lambda_2^*, a_2^*) \in \hat{\Gamma}_2(\mu^*)$, where $\mu^{*i} = \bar{\lambda}_1^* + \sum_{\hat{e}} \lambda_1^{\hat{e}*}[\pi_i(e_1^*) - \pi_i(\hat{e})]/\pi_i(e_1^*)$.*

Proof. See Appendix. □

In combination Propositions 9 and 14 imply that if the sequential Lagrangian admits a saddle, then the recursive dual approach delivers the optimal primal value and necessary conditions for optimal primal solutions.

[19] The derivation of the recursive primal involves a similar interchange of a conditional sup and an inf operation. This *is* immediately possible in the current setting.

We conclude with a duality result that parallels Proposition 12. We say that the dual continuation problem (59) admits a saddle at μ^i if its objective:

$$G(a_2^i, \lambda_2^i; \mu^i) = f(a_2^i) + \sum_{j=1}^{N} \pi_j(e_2) \left[\mu^i u\left(w_2^{ij}, e_2\right) + \sum_{\hat{e}\in E} \lambda_2^{\hat{e},i} \left(u(w_2^{ij}, e_2^i) - \frac{\pi_j(\hat{e})}{\pi_j(e_2^i)} u(w_2^{ij}, \hat{e}) \right) \right]$$

has a saddle point on $\widetilde{A} \times \Lambda_2$. The definition of a saddle for the primal continuation problem (57) at y^i is analogous to before. Following the steps in the proofs of Propositions 11 and 12 for the special case in which ϕ and x are absent gives the following result.

Proposition 15. *Suppose (a^*, λ^*) is a saddle point of the sequential Lagrangian (56) and let $\mu^{*i} = \bar{\lambda}_1^* + \sum_{\hat{e}} \lambda_1^{\hat{e}*}[\pi_i(e_1^*) - \pi_i(\hat{e})]/\pi_i(e_1^*)$ and $y^{*i} = m(a_2^{*i})$.*

i) Both the primal continuation problem (57) at y^{*i} and the dual continuation problem (59) at μ^{*i} admit a saddle.

ii) The value functions of the primal and dual continuation problems satisfy:

$$W_2(\mu^{*i}) = \sup_{Y} \; V_2(y^i) + \mu^{*i} \cdot y^i = (-V_2)^\star(\mu^{*i}) \tag{60}$$

$$-V_2(y^{*i}) = \sup_{M} \mu^i \cdot y^{*i} - W_2(\mu^i) = (W_2)^\star(y^{*i}). \tag{61}$$

iii) W_2 and $-V_2$ are sub-differentiable at μ^{*i} and y^{*i}, respectively with: $\mu^{*i} \in \partial[-V_2(y^{*i})]$ and $y^{*i} \in \partial W_2(\mu^{*i})$.

Proof. See Appendix. □

The results of this section rely on one central insight: if the underlying sequential optimization problem admits a saddle then the order in which inf and sup operations are applied is immaterial for the value. Moreover, the saddle is a solution of the problem for any order of these operations. Aside from saddle point existence, the only feature of the moral hazard framework that we have exploited in this section is the ability to meaningful rearrange the Lagrangian. Consequently, our results extend to any non-additively separable framework with this feature.

7 Comparison with the Recursive Saddle Point Method

Marcet and Marimon (1999) propose a different method from ours. Instead of recursively decomposing sup-inf or inf-sup operations, they recursively calculate the saddle points of a Lagrangian. In addition, instead of using only primal or only dual states as we do, they use only backwards-looking states. Thus, while we mix the direction of state variables, they mix primal and dual states.

An advantage of our approach is that the derivation of the recursive dual requires much weaker conditions than that of the recursive saddle point problem. Section 3.3 shows that an equivalence holds between the sequential and recursive dual problems in additively separable frameworks. This equivalence is independent of any explicit saddle assumptions. In particular, it does not require that the sequential Lagrangian admits a saddle. Such an assumption is only required to link the value and solutions of the (recursive) dual to the sequential primal. In contrast for the recursive saddle point method (RSPM) to be well defined, it is necessary not only that the sequential optimization method admits a saddle, but also that every continuation problem starting from an arbitrarily chosen vector of (primal and dual) states admits a saddle as well. We illustrate this important difference between our approach and the RSPM with the following example.

Example 3 (Recursive Saddle Points)**.** Consider the two period problem:

$$\max_{(a_1,a_2)\in[0,2]^2} \quad -a_1 - a_2 \tag{62}$$

$$\text{s.t.} \quad u(a_1) + u(a_2) \geq 2 \quad \text{and} \quad u(a_2) \geq 1/4, \tag{63}$$

where

$$u(a) = \begin{cases} a^2 & \text{if } a \leq 1 \\ \sqrt{a} & \text{else.} \end{cases}$$

Notice that u is strictly convex for $a < 1$ and strictly concave for $a > 1$. It is not difficult to verify that this problem's optimal value is $V^* = -2$ and its unique solution is: $a^* = (a_1^*, a_2^*) = (1, 1)$. Substituting a^* into the sequential Lagrangian,

$$\mathscr{L}(a, \lambda) = -a_1 + a_2 + \lambda_1(u(a_1) + u(a_2) - 2) + \lambda_2(u(a_2) - 1/4),$$

gives an expression that is constant in λ_1 and strictly increasing in λ_2. Thus $\lambda^* =$

$(\lambda_1^*, \lambda_2^*) = (1, 0)$ minimizes $\mathscr{L}(a^*, \lambda)$. On the other hand, it is easy to verify that a^* maximizes $\mathscr{L}(a, \lambda^*)$. Thus (a^*, λ^*) is a saddle point of $\mathscr{L}$.

Now, for $g : C \times D \to \mathbb{R}$, let $\text{saddle}_{C|D}\, g(c, d)$ denote the saddle value operation: $\text{saddle}_{C|D}\, g(c, d) = g(c^*, d^*)$, $c^* \in \text{argmax}_C\, g(c, d^*)$, $d^* \in \text{argmin}_D\, g(c^*, d)$. The recursive version of $\text{saddle}_{A|\Lambda}\, \mathscr{L}(a, \lambda)$ is given by:

$$\begin{aligned} \tilde{W}_1 \quad &:= \underset{A_1|\Lambda_1}{\text{saddle}} - a_1 + \lambda_1(u(a_1) - 2) + \tilde{W}_2(\mu_2) \\ &\quad \text{s.t. } \mu_2 = \lambda_1, \end{aligned} \tag{64}$$

where:

$$\tilde{W}_2(\mu_2) := \underset{A_2|\Lambda_2}{\text{saddle}} - a_2 + \mu_2 u(a_2) + \lambda_2(u(a_2) - 1/4). \tag{65}$$

Observe that the first period problem (64) is well defined only if the value of the second period problem (65), $\tilde{W}_2$, is specified for every $\mu_2 \geq 0$. But it is not difficult to see that (65) does not admit a solution when $\mu_2 < 1$. Consider for instance the case $\mu_2 = 0$. If $a_2 < 1/2$, then the objective function of the second period problem is strictly decreasing in λ_2 and thus no finite minimizer can exist. If instead $a_2 > 1/2$, then the objective is strictly increasing in λ_2, meaning that $\lambda_2 = 0$ is the unique minimizer. But for $\lambda_2 = 0$ the maximizing value of a_2 is 0. Finally, observe that for every $\lambda_2 > 0$ and every $a_2 < 1$ the objective function is strictly convex in a_2. Thus, there can be no λ_2 such that the objective is maximized at an a_2 in the interior of $[0, 1]$.

So this example is an instance where the RSPM cannot be applied, because it violates the condition that the continuation problem admits a saddle for each value of the state variable.[20] On the other hand, there are no problems in applying the dual recursive method (DRM) to this example. Since the sequential Lagrangian admits a saddle we know that the dual recursive approach delivers the correct value and that it also yields necessary conditions for the optimal solutions.

A second difference between our approach and the RSPM proposed by Marcet and Marimon (2011), lies in the fact that we obtain necessity under far weaker assumptions. In particular, Marcet and Marimon (2011) prove that the policy of the RSPM generates every solution of the underlying optimization problem under the assumption that there is only one such solution. Our necessity result requires no uniqueness assumptions

[20] This fact does not hinge on the exact functional form of the agent's utility. The crucial feature of u is that it is not concave everywhere but instead is convex over parts of its domain.

whatsoever.

Marcet and Marimon (2011) show that their method yields necessary and sufficient conditions for an optimum if the functions that define the per period returns and constraints satisfy (strict) concavity. The concavity assumptions that we have imposed in Section 3.3 in order to prove our sufficiency result are the exact analogue of their assumptions for our framework.[21] So while we get necessity under weaker assumptions than Marcet and Marimon (2011), we do not need to impose stronger assumptions in order to obtain sufficiency.

[21]See Assumptions A6-A6s in Marcet and Marimon (2011).

Appendix A: Proofs

Proof of Proposition 6. Let Gr $X = \{(a_1, x) | a_1 \in A_1, x = l(a_1)\}$ and $A_2(y) := \{a_2 \in A_2 : m(a_2) = y\}$. Then:

$$\begin{aligned}
V^* &= \sup_{A} \inf_{\Lambda} \{f_1(a_1) + f_2(a_2) + \lambda_1[g_1(a_1) + c_1 \cdot m(a_2)] + \lambda_2[c_2 \cdot l(a_1) + g_2(a_2)]\} \\
&= \sup_{\text{Gr } X \times Y} \sup_{A_2(y)} \inf_{\Lambda} \{f_1(a_1) + f_2(a_2) + \lambda_1[g_1(a_1) + c_1 \cdot y] + \lambda_2[c_2 \cdot x + g_2(a_2)]\} \\
&= \sup_{\text{Gr } X \times Y} \sup_{A_2(y)} \inf_{\Lambda_1} \left\{ f_1(a_1) + \lambda_1[g_1(a_1) + c_1 \cdot y] + \inf_{\Lambda_2} \{f_2(a_2) + \lambda_2[c_2 \cdot x + g_2(a_2)]\} \right\} \\
&= \sup_{\text{Gr } X \times Y} \inf_{\Lambda_1} \left\{ f_1(a_1) + \lambda_1[g_1(a_1) + c_1 \cdot y] + \sup_{A_2(y)} \inf_{\Lambda_2} \{f_2(a_2) + \lambda_2[c_2 \cdot x + g_2(a_2)]\} \right\} \\
&= \sup_{\text{Gr } X \times Y} \inf_{\Lambda_1} \{f_1(a_1) + \lambda_1[g_1(a_1) + c_1 \cdot y] + V_2(x, y)\} = V_1,
\end{aligned}$$

where the first line follows from the definition of V^*, the second from the definitions of Gr X, x and y and the third from the decomposition properties of the infimum operation. The fourth interchanges the conditional supremum operation over a_2 with the infimum operation over λ_1. It is admissible because the choice of λ_1 is independent of a_2 given y. The fifth and sixth equalities stem from the definitions of V_2 and V_1 respectively.

Let $a^* = (a_1^*, a_2^*)$ and $\lambda^* = (\lambda_1^*, \lambda_2^*)$ be as in the statement of the proposition and let $x^* = l(a_1^*)$ and $y^* = m(a_2^*)$. Then:

$$\lambda^* \in \operatorname*{argmin}_{\Lambda} f_1(a_1^*) + f_2(a_2^*) + \lambda_1[g_1(a_1^*) + c_1 \cdot y^*] + \lambda_2[c_2 \cdot x^* + g_2(a_2^*)].$$

Decomposing this and adding the constant term $V_2(x^*, y^*)$ gives:

$$\begin{aligned}
\lambda_1^* &\in \operatorname*{argmin}_{\Lambda_1} f_1(a_1^*) + \lambda_1[g_1(a_1^*) + c_1 \cdot y^*] + V_2(x^*, y^*) \\
\lambda_2^* &\in \operatorname*{argmin}_{\Lambda_2} f_2(a_2^*) + \lambda_2[c_2 \cdot x^* + g_2(a_2^*)].
\end{aligned}$$

In particular, since a^* is feasible and $g(a^*) \geq 0$, $\lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] = 0$ and $\lambda_2^*[c_2 \cdot x^* +$

$g_2(a_2^*)] = 0$. Also,

$$
\begin{aligned}
V_1 &= V^* \\
&= f_1(a_1^*) + f_2(a_2^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + \lambda_2^*[c_2 \cdot x^* + g_2(a_2^*)] \\
&= f_1(a_1^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + f(a_2^*) + \lambda_2^*[c_2 \cdot x^* + g_2(a_2^*)] \\
&= f_1(a_1^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + \inf_{\Lambda_2}\{f(a_2^*) + \lambda_2[c_2 \cdot x^* + g_2(a_2^*)]\} \\
&\leq f_1(a_1^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + \sup_{A_2(y^*)} \inf_{\Lambda_2} \{f(a_2) + \lambda_2[c_2 \cdot x^* + g_2(a_2)]\} \\
&= f_1(a_1^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + V_2(x^*, y^*) \\
&= \inf_{\Lambda_1} f_1(a_1^*) + \lambda_1[g_1(a_1^*) + c_1 \cdot y^*] + V_2(x^*, y^*) \\
&\leq \sup_{\mathrm{Gr}\, X \times Y} \inf_{\Lambda_1} f_1(a_1) + \lambda_1[g_1(a_1) + c_1 \cdot y] + V_2(x, y) = V_1,
\end{aligned}
$$

where the first line is from the first part of the proposition, the second is from the definitions of V^*, a^* and λ^*, the third is a simple rearrangement, the fourth from the previously proved property of λ_2^*, the fifth from the fact that $a_2^* \in A_2(y^*)$, the sixth from the definition of V_2, the seventh from the previously proved property of λ_1^*, the eighth from the fact that $(a_1^*, x^*, y^*) \in \mathrm{Gr}X \times Y$ and the definition of V_1. Thus, we have (using the sixth line above): $V_1 = f_1(a_1^*) + \lambda_1^*[g_1(a_1^*) + c_1 \cdot y^*] + V_2(x^*, y^*)$. Combining this with the previously proved first period complementary slackness condition gives $V_1^* = f_1(a_1^*) + V_2(x^*, y^*)$. Similarly, we have (using the third and sixth lines above) $V_2(x^*, y^*) = f_2(a_2^*) + \lambda_2^*[c_2 \cdot x^* + g_2(a_1^*)]$. Combining this with the previously proved second period complementary slackness condition gives $V_2(x^*, y^*) = f_2(a_2^*)$. Putting the pieces together, we conclude that $(a_1^*, x^*, y^*, \lambda_1^*) \in \Gamma_1$ and $(a_2^*, \lambda_2^*) \in \Gamma_2(x^*, y^*)$. The converse is proved in a similar fashion. □

Proof of Proposition 10. *(i)* Follows from Borwein and Lewis (2006), Theorem 3.2.8, p. 44. *(ii)* There is at most one solution to the sequential primal problem since the constraint set is convex ($g_1 + c_1 \cdot m$ and $c_2 \cdot l + g_2$ are both concave and thus also quasi-concave) and the objective $f_1 + f_2$ is strictly quasi-concave (strict concavity of f_1 and f_2 implies strict quasi-concavity of $f_1 + f_2$). Since A^* is non-empty by assumption, it follows that $A^* = \{a^*\}$ for some $a^* \in A$. By Proposition 7, $W_1 = W^*$, and by Proposition 8, A_{IS} coincides with the set of control sequences generated by the recursive dual policy correspondences. It remains to show that $A_{IS} = A^*$.

By Proposition 1 and part i) of this proposition, there is a $\lambda^* \in \Lambda_{IS}$ such that (a^*, λ^*) is a saddle point. But then, from the definition of a saddle point, $a^* \in A_{IS}$ and $A^* \subset A_{IS}$. Suppose that there is an $\tilde{a} \in A_{IS} \backslash A^*$. Since $\tilde{a} \in A_{IS}$, there is a $\tilde{\lambda} \in \Lambda_{IS}$ such that $\tilde{a}$ maximizes $\mathscr{L}(\cdot, \tilde{\lambda})$. In fact since $\mathscr{L}(\cdot, \tilde{\lambda})$ is strictly concave, $\tilde{a}$ is the unique maximizer of $\mathscr{L}(\cdot, \tilde{\lambda})$. On the other hand, since the set of saddle points of the Lagrangian coincides with $A^* \times \Lambda_{IS}$, $(a^*, \tilde{\lambda})$ is a saddle point and a^* a maximizer of $\mathscr{L}(\cdot, \tilde{\lambda})$. This contradiction implies that $A_{IS} \backslash A^* = \emptyset$ and so a^* is the only element of A_{IS}. □

Proof of Proposition 11. *(i)* The conjugate of $-V_2$ is given by:

$$
\begin{aligned}
(-V_2)^\star(\mu, -\phi) &:= \sup_{X \times Y} \; -\phi \cdot x + \mu \cdot y + V_2\,(x, y) \\
&= \sup_{X \times Y} \; -\phi \cdot x + \mu \cdot y + \sup_{A_2} \inf_{\Lambda_2} \inf_{\Delta} \; f_2\,(a_2) + \lambda_2 \cdot [g_2\,(a_2) + c_2 \cdot x] + \delta\,[y - m\,(a_2)] \\
&= \sup_{X \times Y} \; \sup_{A_2} \inf_{\Lambda_2} \inf_{\Delta} -\phi \cdot x + \mu \cdot y + \; f_2\,(a_2) + \lambda_2 \cdot [g_2\,(a_2) + c_2 \cdot x] + \delta\,[y - m\,(a_2)]\,.
\end{aligned}
$$

Obviously, y and a_2 will be chosen so that $y = m\,(a_2)$, and so:

$$
(-V_2)^\star\,(\mu, -\phi) = \sup_{X \times A_2} \inf_{\Lambda_2} -\phi \cdot x + \mu \cdot m\,(a_2) + \; f_2\,(a_2) + \lambda_2 \cdot g_2\,(a_2) + \lambda_2 \cdot c_2 \cdot x. \tag{66}
$$

On the other hand:

$$
\begin{aligned}
W_2\,(\mu, \phi) &:= \inf_{\Lambda_2} \sup_{A_2} \sup_{Z} \; f_2\,(a_2) + \mu \cdot m\,(a_2) + \lambda_2 \cdot g_2\,(a_2) + z\,[\phi - \lambda_2 \cdot c_2] \\
&\geq \inf_{\Lambda_2} \sup_{A_2 \times X} \; f_2\,(a_2) + \mu \cdot m\,(a_2) + \lambda_2 \cdot g_2\,(a_2) + \lambda_2 \cdot c_2 \cdot x - \phi \cdot x,
\end{aligned} \tag{67}
$$

where in the second line the change of variable $z = -x$ is made. The inequality sign follows from the fact that the set from which x is chosen X is a subset of $-Z$. Hence, (66), (67) and weak duality imply that $W_2 \geq (-V_2)^\star$. The second inequality (31) in Proposition 11 (i) is proved analogously by comparing $W_2^\star$ to $-V_2$. Part (ii) of the proposition is immediate from the definition of a saddle point and Proposition 1. □

Proof of Proposition 12. (i) We first show that the dual continuation problem (27) admits a saddle point at $(\mu^*, \phi^*) = (\lambda_1^* \cdot c_1, \lambda_2^* \cdot c_2)$. Since (a^*, λ^*) is a saddle of $\mathscr{L}(a, \lambda)$, it follows that for all $a_2 \in A_2$ and $\lambda_2 \in \Lambda_2$:

$$
\mathscr{L}(a^*, \lambda_1^*, \lambda_2) \geq \mathscr{L}(a^*, \lambda^*) \geq \mathscr{L}(a_1^*, a_2, \lambda^*).
$$

Thus, (a_2^*, λ_2^*) is a saddle of $f_2(a_2) + \mu^* \cdot m(a_2) + \lambda_2 \cdot g_2(a_2) + (\lambda_2 \cdot c_2 - \phi^*) \cdot x^*$. Hence, for all $(\lambda_2; a_2, z) \in \Lambda_2 \times A_2 \times \mathbb{R}^{r_2}$,

$$\begin{aligned} f_2(a_2) + \mu^* m(a_2) + \lambda_2^* \cdot g_2(a_2) + (\lambda_2^* \cdot c_2 - \phi^*) \cdot z \\ \leq f_2(a_2^*) + \mu^* m(a_2) + \lambda_2^* \cdot g_2(a_2) + (\lambda_2^* \cdot c_2 - \phi^*) \cdot x^* \\ \leq f_2(a_2^*) + \mu^* m(a_2) + \lambda_2 \cdot g_2(a_2) + (\lambda_2 \cdot c_2 - \phi^*) \cdot x^*, \end{aligned}$$

where the first inequality uses the fact that a_2^* is maximal for $f_2(a_2) + \mu^* m(a_2) + \lambda_2^* \cdot g_2(a_2)$ given the saddle property of (a_2^*, λ_2^*) and the fact that $\lambda_2^* \cdot c_2 - \phi^* = 0$ and the second uses the minimality of λ_2^* given the saddle property of (a_2^*, λ_2^*). Hence, $(\lambda_2^*; a_2^*, x^*)$ is a saddle for the continuation dual problem at (μ^*, ϕ^*). The proof that the continuation primal problem admits a saddle at (x^*, y^*) is analogous.

(ii) The proof is done in two steps. First, we show the following double implication:

$$\begin{bmatrix} -\phi^* \\ \mu^* \end{bmatrix} \in \partial\left[-V_2\left(x^*, y^*\right)\right] \Leftrightarrow \begin{bmatrix} y^* \\ -x^* \end{bmatrix} \in \partial W_2\left(\mu^*, \phi^*\right) \tag{68}$$

assuming sub-differentiability of the functions. Then we show that $-V_2$ is sub-differentiable at (x^*, y^*), with $(-\phi^*, \mu^*)$ a sub-gradient.

Step 1: [$\Leftarrow$ of (68)] Let $(y^*, -x^*) \in \partial W_2(\mu^*, \phi^*)$. By Proposition 2, we have $(\mu^*, \phi^*) \in \partial W_2^\star(y^*, -x^*)$. From the definition of a sub-differential, Proposition 11 and (i) of this proposition:

$$\begin{aligned} -V_2(x, y) \geq W_2^\star(y, -x) \geq \quad & W_2^\star(x^*, y^*) - \phi^* \cdot [x - x^*] + \mu^* \cdot [y - y^*] \\ = & -V_2(x^*, y^*) - \phi^* \cdot [x - x^*] + \mu^* \cdot [y - y^*]. \end{aligned}$$

So that $(-\phi^*, \mu^*) \in \partial[-V_2(x^*, y^*)]$ as desired.

[$\Rightarrow$ of (68)] Let $(-\phi^*, \mu^*) \in \partial[-V_2(x^*, y^*)]$. Again using Proposition 2, we have $(x^*, y^*) \in \partial(-V_2)^\star(-\phi^*, \mu^*)$. This together with the definition of a sub-differential, Proposition 11 and (i) of this proposition gives:

$$W_2(\mu, \phi) \geq (-V_2)^\star(\mu, -\phi) \geq W_2(\mu^*, \phi^*) + y^* \cdot (\mu - \mu^*) - x^* \cdot (\phi - \phi^*)$$

which implies that $(y^*, -x^*) \in \partial W_2(\mu^*, \phi^*)$ as desired.

Step 2: From Proposition 11,

$$-V_2(x^*, y^*) = W_2^\star(-x^*, y^*) := \sup_{F \times M} -W_2(\mu, \phi) + \mu \cdot y^* - \phi \cdot x^*. \tag{69}$$

We now show that (ϕ^*, μ^*) solves the previous maximization problem. We have:

$$
\begin{aligned}
-V_2(x^*, y^*) &\geq -W_2(\mu^*, \phi^*) + \mu^* \cdot y^* - \phi \cdot x^* \\
&= -\inf_{\Lambda_2} \sup_{A_2 \times \mathbb{R}^{r_2}} \left[f(a_2) + \mu^* \cdot m(a_2) + \lambda_2 \cdot g(a_2) + x \cdot (\phi^* - \lambda_2 \cdot c_2) \right] + \mu^* \cdot y^* - \phi^* \cdot x^* \\
&\geq - \sup_{A_2 \times \mathbb{R}^{r_2}} \left[f(a_2) + \mu^* \cdot m(a_2) + \lambda_2^* \cdot g(a_2) + x \cdot (\phi^* - \lambda_2^* \cdot c_2) \right] + \mu^* \cdot y^* - \phi^* \cdot x^* \\
&= -f(a_2^*) - \mu^* \cdot m(a_2^*) - \lambda_2^* \cdot g(a_2^*) + \mu^* \cdot y^* - \phi^* \cdot x^* \\
&= -f(a_2^*) - \lambda_2^* \cdot g(a_2^*) - \lambda_2^* \cdot c_2 \cdot x^* = -V_2(x^*, y^*),
\end{aligned}
$$

where the first line uses (69), the second the definition of W_2, the third the fact that λ_2^* need not be maximal, the fourth the equality $\phi^* = \lambda_2^* \cdot c_2$ and the saddle point property of a_2^*, the fifth $y^* = m(a_2^*)$ and the definition of V_2. We hence have shown that:

$$
-W_2(\mu^*, \phi^*) + \mu^* \cdot y^* - \phi^* \cdot x^* \geq -W_2(\mu, \phi) + \mu \cdot y^* - \phi \cdot x^*.
$$

A simple rearrangement of terms implies that:

$$
W_2(\mu, \phi) \geq W_2(\mu^*, \phi^*) + (\mu - \mu^*) \cdot y^* - (\phi - \phi^*) \cdot x^*.
$$

which is the definition of a sub-gradient. So, $(-x^*, y^*) \in \partial W_2(\phi^*, \mu^*)$, and we are finished. □

Proof of Proposition 13. Part (i): Observe that the following sequence of (in-)equalities holds

$$
\begin{aligned}
\hat{W}^*(\bar{\mu}_1, \bar{x}_1) &\overset{(1)}{=} W_1(\bar{\mu}_1, \phi_1^*) + \phi_1^* \bar{x}_1 \\
&\overset{(2)}{=} \inf_{\Lambda} \sup_{A,Z} \mathscr{L}^D(a, \lambda, z; \bar{\mu}_1, \phi_1^*) + \phi_1^* \bar{x}_1 \\
&\overset{(3)}{\geq} \inf_{\Lambda} \sup_{A} \mathscr{L}^D(a, \lambda, \bar{x}_1; \bar{\mu}_1, \phi_1^*) + \phi_1^* \bar{x}_1 \\
&\overset{(4)}{=} \inf_{\Lambda} \sup_{A} H(a, \lambda) + \bar{\mu}_1 \cdot \left(\sum_{\tau=1}^{2} m(a_\tau) \right) + \left(\sum_{\tau=1}^{2} \lambda_\tau b_{1,\tau} \right) \cdot \bar{x}_1 \\
&\overset{(5)}{\geq} \sup_{A} \inf_{\Lambda} H(a, \lambda) + \bar{\mu}_1 \cdot \left(\sum_{\tau=1}^{2} m(a_\tau) \right) + \left(\sum_{\tau=1}^{2} \lambda_\tau b_{1,\tau} \right) \cdot \bar{x}_1
\end{aligned}
$$

$$\overset{(6)}{=} \sup_A \inf_\Lambda H(a,\lambda) + \bar{\mu}_1 \cdot \Big(\sum_{\tau=1}^{2} m(a_\tau) - y^* \Big) + \bar{\mu}_1 \cdot y_1^* + \Big(\sum_{\tau=1}^{2} \lambda_\tau b_{1,\tau} \Big) \cdot \bar{x}_1$$
$$\overset{(7)}{=} \sup_A \inf_\Lambda \mathscr{L}^P(a,\lambda,\bar{\mu}_1;\bar{x}_1,y_1^*) + \bar{\mu}_1 \cdot y_1^*$$
$$\overset{(8)}{\geq} \sup_A \inf_{\Lambda,\Gamma} \mathscr{L}^P(a,\lambda,\gamma;\bar{x}_1,y_1^*) + \bar{\mu}_1 \cdot y_1^*$$
$$\overset{(9)}{=} V_1(\bar{x}_1,y_1^*) + \bar{\mu}_1 \cdot y_1^*$$
$$\overset{(10)}{=} \hat{V}^*(\bar{\mu}_1,\bar{x}_1).$$

The first equality sign is definitional. The second follows from the fact that $W_1(\bar{\mu}_1,\phi^*)$ is equal to the value of the sequential dual starting at $(\bar{\mu}_1,\phi_1^*)$ (i.e. $W_1(\bar{\mu}_1,\phi^*)$ equals $\inf_\Lambda \sup_{A,Z} \mathscr{L}^D(a,z,\lambda;\bar{\mu}_1,\phi_1^*)$). The inequality in line three is implied by the fact that instead of calculating the supremum with respect to z we evaluate $\mathscr{L}^D$ at $z=\bar{x}_1$. The next equality follows immediately from the definition of $\mathscr{L}^D$. The inversion of the supremum and infimum operators implies the inequality sign in line five. Line six is obtained by adding and subtracting $\bar{\mu}_1 \cdot y_1^*$ to the previous line. The equality in line 7 follows from the definition of $\mathscr{L}^P$. The inequality in the next line is implied by the fact that we introduce an infimum operation with respect to a variable that was held constant in line 7. Line nine follows from the fact that $\sup_A \inf_{\Lambda,\Gamma} \mathscr{L}^P(a,\lambda,\gamma;\bar{x}_1,y_1^*)$ and $V_1(\bar{x}_1,y_1^*)$ coincide. The equality in the last line is again definitional.

Part (ii): Assume first that $\hat{W}^*(\bar{\mu}_1,\bar{x}_1) = \hat{V}^*(\bar{\mu}_1,\bar{x}_1)$. From this equation it follows that all inequality signs in the above expression must hold as equalities. Moreover, it is also straightforward to see that if λ^* solves the problem in line 2 above then it must be a solution of the problem in line 3 as well. Specifically,

$$\begin{aligned}
\sup_A \mathscr{L}^D(a,\lambda^*,\bar{x}_1;\bar{\mu}_1,\phi_1^*) + \phi_1^*\bar{x}_1 &\overset{(i)}{\geq} \inf_\Lambda \sup_A \mathscr{L}^D(a,\lambda,\bar{x}_1;\bar{\mu}_1,\phi_1^*) + \phi_1^*\bar{x}_1 \\
&\overset{(ii)}{=} \inf_\Lambda \sup_{A,Z} \mathscr{L}^D(a,\lambda,z;\bar{\mu}_1,\phi_1^*) + \phi_1^*\bar{x}_1 \\
&\overset{(iii)}{=} \sup_{A,Z} \mathscr{L}^D(a,\lambda^*,z;\bar{\mu}_1,\phi_1^*) + \phi_1^*\bar{x}_1 \\
&\overset{(iv)}{\geq} \sup_A \mathscr{L}^D(a,\lambda^*,\bar{x}_1;\bar{\mu}_1,\phi_1^*) + \phi_1^*\bar{x}_1,
\end{aligned}$$

where (i) follows from the fact that λ^* need not attain the infimum on the right hand side, (ii) follows from the equality of lines 2 and 3 above, (iii) is implied by the fact λ^* attains the minimum in the second line and (iv) follows from the fact that $\bar{x}_1$ need not attain the supremum in the Lagrangian. Hence, (i) must hold with equality and λ^* is minimal in line 3 above. An analogous argument applies to the problems in line 7 and 8. That is, a^*, which solves the problem in line 8, is also a solution to the problem in line 7.

Now observe that the objective functions in lines four and five coincide and that both problems have a solution (respectively, λ^* and a^*). But this means that (a^*, λ^*) must be a saddle of

$$G(a, \lambda; \bar{\mu}_1, \bar{x}_1) := H(a, \lambda) + \bar{\mu}_1 \cdot \Big(\sum_{\tau=1}^{2} m(a_\tau) \Big) + \Big(\sum_{\tau=1}^{2} \lambda_\tau b_{1,\tau} \Big) \cdot \bar{x}_1.$$

We now verify that $(a^*, \lambda^*, \bar{\mu}_1)$ is a saddle point of $\mathscr{L}^P(a, \lambda, \gamma; \bar{x}_1, y_1^*)$. Recall that a^* solves the problem in line 8 above. Infinite penalization implies that $\sum_{\tau=1}^{2} m(a_\tau) = y_1^*$. Thus, $\mathscr{L}^P(a^*, \lambda, \gamma; \bar{x}_1, y_1^*)$ is constant in γ and so it is minimized by $\gamma = \bar{\mu}_1$. Moreover, $\mathscr{L}^P(a^*, \lambda, \bar{\mu}_1; \bar{x}_1, y_1^*)$ varies with λ in the same way as $G(a^*, \lambda; \bar{\mu}_1, \bar{x}_1)$ (the two functions differ only by an additive term that does not depend on λ). Since (a^*, λ^*) is a saddle of $G(a, \lambda; \bar{\mu}_1, \bar{x}_1)$ we can conclude that λ^* must be a minimizer of $\mathscr{L}^P(a^*, \lambda, \bar{\mu}_1; \bar{x}_1, y_1^*)$. Conversely, fix $(\lambda, \gamma) = (\lambda^*, \bar{\mu}_1)$. The resulting function $\mathscr{L}^P(a, \lambda^*, \bar{\mu}_1; \bar{x}_1, y_1^*)$ differs from $G(a, \lambda^*; \bar{\mu}_1, \bar{x}_1)$ only by a constant. Thus, since the latter function is maximized by a^*, so must $\mathscr{L}^P(a, \lambda^*, \bar{\mu}_1; \bar{x}_1, y_1^*)$.

The arguments required to show that $(a^*, \lambda^*, \bar{x}_1)$ is a saddle point of $\mathscr{L}^D(a, \lambda, z; \bar{\mu}_1, \phi_1^*)$ are analogous to those just employed for the primal Lagrangian case. We therefore omit the details.

For the converse, notice that if $(a^*, \lambda^*, \bar{\mu}_1)$ and $(a^*, \lambda^*, \bar{x}_1)$ are saddle points of $\mathscr{L}^P(\cdot; \bar{x}_1, y_1^*)$ and $\mathscr{L}^D(\cdot; \bar{\mu}_1, \phi_1^*)$, respectively, then the inequalities in lines (3), (5) and (8) hold as equalities. This immediately gives the desired result, $\hat{V}^*(\bar{\mu}_1, \bar{x}_1) = \hat{W}^*(\bar{\mu}_1, \bar{x}_1)$. iii) The proof is analogous to Propositions 11 and 12 and is omitted.

□

Proof of Proposition 14. *(i)* Notice that:

$$W^* = \inf_{\Lambda_1 \times \Lambda_2} \sup_{A_1 \times A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2)$$

and

$$W_1 = \inf_{\Lambda_1} \sup_{A_1} \inf_{\Lambda_2} \sup_{A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2).$$

Since:

$$\inf_{\Lambda_2} \sup_{A_1} \sup_{A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2) \geq \sup_{A_1} \inf_{\Lambda_2} \sup_{A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2),$$

it follows that $W^* \geq W_1$. But the same argument also implies that:

$$W_1 \geq \inf_{\Lambda_1} \sup_{A_1} \sup_{A_2} \inf_{\Lambda_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2) \geq \sup_{A_1} \sup_{A_2} \inf_{\Lambda_1} \inf_{\Lambda_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2) = V^*.$$

If a saddle exists, then $W^* \geq W_1 \geq V^* = W^*$ and so $W^* = W_1$.

(ii) If (a^*, λ^*) is a sequential saddle, then (a_2^*, λ_2^*) is a saddle of $\mathscr{L}(a_1^*, a_2, \lambda_1^*, \lambda_2)$. Hence, by Proposition 1, $(\lambda_2^*, a_2^*) \in \hat{\Gamma}_2(\mu^*)$, where μ^* is the value of the state implied by (a_1^*, λ_1^*). Since there exists a saddle the order of the inf and sup operations does not matter. In particular,

$$\inf_{\Lambda_1} \sup_{A_1} \left[\inf_{\Lambda_2} \sup_{A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2) \right] = \sup_{A_1} \inf_{\Lambda_1} \left[\inf_{\Lambda_2} \sup_{A_2} \mathscr{L}(a_1, a_2, \lambda_1, \lambda_2) \right] = \mathscr{L}(a_1^*, a_2^*, \lambda_1^*, \lambda_2^*).$$

But then, by Proposition 1, (λ_1^*, a_1^*) must belong to $\hat{\Gamma}_1$.

□

Proof of Proposition 15. The first statement follows from the proof of Proposition 12 (i). When μ^* is given and the state ϕ is absent, the lines of proof in Propositions 11 (ii) and 12 (ii) can be used, step by step, to show results (ii) and (iii) above. □

References

Acemoğlu, D., M. Golosov, and A. Tsyvinski (2010). Dynamic mirrlees taxation under political economy constraints. *Review of Economic Studies 77*, 841–881.

Aiyagari, R., A. Marcet, T. Sargent, and J. Seppälä (2002). Optimal taxation without state-contingent debt. *Journal of Political Economy 110*, 1220–1254.

Bellman, R. E. (1957). *Dynamic Programing*. Princeton University Press.

Borwein, J. and A. Lewis (2006). *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer.

Chien, Y., H. Cole, and H. Lustig (2011). A multiplier approach to understanding the macro implications of household finance. *Review of Economic Studies 78*, 199–234.

Cole, H. and F. Kubler (2010). Recursive contracts, lotteries and weakly concave pareto sets. University of Pennsylvania.

Kehoe, P. and F. Perri (2002). International business cycles with endogenous incomplete markets. *Econometrica 70*, 907 – 928.

Kydland, F. and E. Prescott (1980). Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control 2*, 79–91.

Marcet, A. and R. Marimon (1999). Recursive contracts. Working paper.

Marcet, A. and R. Marimon (2011). Recursive contracts. Working Paper.

Marimon, R. and V. Quadrini (2006). Competition, innovation and growth with limited commitment. NBER Working Paper 12474.

Mele, A. (2010). Repeated moral hazard and recursive lagrangeans. Working paper.

Messner, M. and N. Pavoni (2004). On the recursive saddle point method. IGIER Working Paper 255.

Messner, M., N. Pavoni, and C. Sleet (2011). On the dual approach to recursive optimization. Working paper.

Rockafellar, T. (1970). *Convex Analysis*. Princeton University Press.

Stokey, N., R. Lucas, and E. Prescott (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.